

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2004, 09:34:43 ; Search time 142 Seconds
(without alignments)
2108.639 Million cell updates/sec

Title: US-09-671-687A-3
Perfect score: 5034
Sequence: 1 MSGSLWSQKVTSPYWEERI.....RLLCDAYMCWQSPYSLYK 949

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 segs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : \$PTREMBL_25.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriaph.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4983	99.0	960	094934	Q94934 homo sapien
2	4971.5	98.8	956	Q9NQC7	Q9nqc7 homo sapien
3	4718.5	93.7	952	Q8CGB0	Q8cgb0 mus musculus
4	4718.5	93.7	959	Q80TQ2	Q80tq2 mus musculus
5	2952	58.6	623	Q8BYL9	Q8byl9 mus musculus
6	2512	49.9	515	Q9NZX9	Q9nzx9 homo sapien
7	1662	33.0	313	Q80VB3	Q80vb3 mus musculus
8	1458.5	29.0	318	Q8BXZ3	Q8bxz3 mus musculus
9	1136.5	22.6	551	Q8IPC3	Q8ipc3 drosophila
10	1136.5	22.6	639	Q8IPC5	Q8ipc5 drosophila
11	1134	22.5	550	Q8IPC4	Q8ipc4 drosophila
12	1132.5	22.5	639	Q8SYF0	Q8syf0 drosophila
13	856	17.0	517	Q9VL04	Q9vl04 drosophila
14	548	10.9	970	Q9U3F9	Q9u3f9 caenorhabdi
15	149.5	3.0	1046	Q55156	Q55156 rattus norv
16	149	3.0	1046	Q9Z0H8	Q9z0h8 mus musculus

17 147.5 2.9 449 4 Q86WU4
18 145 2.9 1556 5 Q9VRP1
19 143.5 2.9 1047 11 Q9EP81
20 143 2.8 547 4 Q8WML1
21 142.5 2.8 429 4 Q96DZ5
22 142 2.8 1012 11 Q7TSI9
23 141 2.8 1012 11 Q8CHU1
24 140.5 2.8 429 4 Q96C99
25 140 2.8 924 16 Q7VDY2
26 139.5 2.8 1082 10 Q94DZ5
27 139.5 2.8 1921 5 Q86BR0
28 138 2.7 724 11 Q8BWD1
29 138 2.7 1024 4 Q14527
30 138 2.7 1391 11 Q922J3
31 137 2.7 868 10 Q9SDN6
32 135.5 2.7 2273 16 Q31152
33 135 2.7 1046 4 Q9UDT6
34 134.5 2.7 657 4 Q7Z3N8
35 132.5 2.6 419 4 Q43611
36 131 2.6 2042 10 Q9M9P8
37 129.5 2.6 832 5 Q8LMZ4
38 128.5 2.6 634 5 Q06279
39 128.5 2.6 840 11 Q8C0S5
40 128 2.5 560 3 Q874X2
41 127 2.5 803 10 Q8H7I2
42 127 2.5 804 10 Q8S7A2
43 127 2.5 804 10 Q7XFN0
44 126.5 2.5 856 5 Q9V4V6
45 125.5 2.5 624 11 Q7TPM8

ALIGNMENTS

RESULT 1
Q94934 PRELIMINARY; PRT; 960 AA.
ID Q94934 Q96EH0

DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein KIAA0849 (Cylindromatosis) (Turban tumor syndrome) (Fragment).
GN KIAA0849
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99156230; PubMed=10048485;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirosewa M., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes XII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";
RL DNA Res. 5:355-364(1998);

RN [2]
RP SEQUENCE OF 8-960 FROM N.A.
RC TISSUE=Uterus;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB020656; BAA74872.2;
DR EMBL; BC012342; AAH12342.1;

DR Genew; HGNC:2584; CYLD.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0004221; F:ubiquitin thiolesterase activity; IEA.
DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
DR InterPro; IPR000938; CAP-Gly.
DR InterPro; IPR001394; Peptidase_C19.
DR Pfam; PF01302; CAP_GLY; 2.
DR Pfam; PF004443; UCH; 1.

```
DR PROSITE; PS0245; CAP GLY 2; 2.
KW PROSITE; PS0235; UCH_2_3; 1.
DR Hypothetical protein.
FT NON_TER
SQ SEQUENCE 960 AA; 107911 MW; 6038410E57A98289 CRC64;

Query Match
Best Local Similarity 99.0%; Score 4983; DB 4; Length 960;
Matches 948; Conservative 0; Mismatches 1; Indels 4; Gaps 4;

QY 1 MSSGLWSQEKVTSPLYWEERIFVLLQECSTVDKQTKLLKVPKSGIGQYIDRSVGHSHRI 60
Db 8 MSSGLWSQEKVTSPLYWEERIFVLLQECSTVDKQTKLLKVPKSGIGQYIDRSVGHSHRI 67
QY 61 PSAKGGKQIGIKLEOPHAVLFVDE-DVVEINEKFTELLAITNCEERFSLFKNNRLS 119
Db 68 PSAKGGKQIGIKLEOPHAVLFVDEKDVVEINEKFTELLAITNCEERFSLFKNNRLS 127
QY 120 KGLQIDVGCPCVKQLRSRGEKFPVVRFRGPIAERTVSGIFFGVLELLEEGRQGGFTDGV 179
Db 128 KGLQIDVGCPCVKQLRSRGEKFPVVRFRGPIAERTVSGIFFGVLELLEEGRQGGFTDGV 187
QY 180 YQKQLFQCEDECG-FVALDKLELIEDDDTALESYAGPGDTMQVELPPELINSRVSLKG 238
Db 188 YQKQLFQCEDECGVFVALDKLELIEDDDTALESYAGPGDTMQVELPPELINSRVSLKV 247
QY 239 GETIESGTIVFCVLPKGSLEGFVGVDMNDPIGNWDGRFDGV-LCSFACVESTILLHIN 297
Db 248 GETIESGTIVFCVLPKGSLEGFVGVDMNDPIGNWDGRFDGVQVLCSPACVESTILLHIN 307
QY 298 DIIPSVTQERRPPPKLAFMRSGVGDGSSSHNPKATGSTSDPGNR-RSELFVTLNGSSV 356
Db 308 DIIPSVTQERRPPPKLAFMRSGVGDGSSSHNPKATGSTSDPGNNRSELFTYLLNGSSV 367
QY 357 DSQPSQSKNTWYIDEVAEDPAKSLTEISTDFDRSSPPLQPPVNSLTATENRPHSLPFL 416
Db 368 DSQPSQSKNTWYIDEVAEDPAKSLTEISTDFDRSSPPLQPPVNSLTATENRPHSLPFL 427
QY 417 TQMPNTNGSIGHSPLSLSAQSVMEELNTAPVQESPPPLAMPNGSHGLEVGSIAEVKENPP 476
Db 428 TQMPNTNGSIGHSPLSLSAQSVMEELNTAPVQESPPPLAMPNGSHGLEVGSIAEVKENPP 487
QY 477 FYGVLRWICQPLNEVLAGELEDEDCAGCTDGTFRGTRYFTCALKKALFVKLSCRDPDS 536
Db 488 FYGVLRWICQPLNEVLAGELEDEDCAGCTDGTFRGTRYFTCALKKALFVKLSCRDPDS 547
QY 537 RFASLPQVSNQIERCNSLAFGGYLSVEVENTPPPKMEKEGLEIMIGKKGIQGHYNSCYL 596
Db 548 RFASLPQVSNQIERCNSLAFGGYLSVEVENTPPPKMEKEGLEIMIGKKGIQGHYNSCYL 607
QY 597 DSTLFCFLAFSSVLDTVLLRPKEKNDVEYYSQTQELLRTTEIVNPLRIYGVYCATKIMKLR 656
Db 608 DSTLFCFLAFSSVLDTVLLRPKEKNDVEYYSQTQELLRTTEIVNPLRIYGVYCATKIMKLR 667
QY 657 KILEKVEAASGTSSEKDPPEELNIFLPHILRVEPLLLKIRSAGQKQVDCYFYQIFMEKNE 716
Db 668 KILEKVEAASGTSSEKDPPEELNIFLPHILRVEPLLLKIRSAGQKQVDCYFYQIFMEKNE 727
QY 717 KYGVPTIQQLLEWSFNSNLKFAEAPSLIIQMPREGKDFKFLFKIIPPSLELNIITDLED 776
Db 728 KYGVPTIQQLLEWSFNSNLKFAEAPSLIIQMPREGKDFKFLFKIIPPSLELNIITDLED 787
QY 777 TPRQCRICGLAMYECCREYDDPDISAGIKIQFCKTQNTQVHLHPKRLNHNKYNPSVLPKD 836
Db 788 TPRQCRICGLAMYECCREYDDPDISAGIKIQFCKTQNTQVHLHPKRLNHNKYNPSVLPKD 847
QY 837 LPDMDWRHGCIFCONNELFAVICIETHSHVAFVKYCKDDSAVLPFDSMADRGQGGFNII 896
Db 848 LPDMDWRHGCIFCONNELFAVICIETHSHVAFVKYCKDDSAVLPFDSMADRGQGGFNII 907
QY 897 PQVTPCPVEGYLKMSELDLSRRIQGCARRLLCDAYMCMYQSPQTMSLYK 949
Db 908 PQVTPCPVEGYLKMSELDLSRRIQGCARRLLCDAYMCMYQSPQTMSLYK 960
```

```
Db 421 FSLTKMNTNGSGHSPSLSAQSVMEELNAPVQESPPPLAMPNGSHGLEVSLAEVKE 480
QY 474 NPPFYGVIRWIGOPPLGNEVLAGELEDEACAGCTDGTFRGTRVFTCALKKALFVKLKSQR 533
Db 481 NPPFYGVIRWIGOPPLGNEVLAGELEDEACAGCTDGTFRGTRVFTCALKKALFVKLKSQR 540
QY 534 PDSRFASLQVSNQIERCNSLAFGGYLVSEVVEENTPPKMEKEGLEIMIGKKKGIOGHYNS 593
Db 541 PDSRFASLQVSNQIERCNSLAFGGYLVSEVVEENTPPKMEKEGLEIMIGKKKGIOGHYNS 600
QY 594 CYLDSTLFCFLAFSSVLDTVLLRPKEKNDVEYVSETOELLARTEIVNPLRIYGYCATKIM 653
Db 601 CYLDSTLFCFLAFSSVLDTVLLRPKEKNDVEYVSETOELLARTEIVNPLRIYGYCATKIM 660
QY 654 KLRKILEKVEAASGFTSEEDKDPPEFLNLFPHILRVEPLKIRSAQKVDQCFYQIFME 713
Db 661 KLRKILEKVEAASGFTSEEDKDPPEFLNLFPHILRVEPLKIRSAQKVDQCFYQIFME 720
QY 714 KNEKVGPTTQQLLEWSFINSNLKFAEAPSLIIQMPRFQKPKLFPKIFPSLELNTDL 773
Db 721 KNEKVGPTTQQLLEWSFINSNLKFAEAPSLIIQMPRFQKPKLFPKIFPSLELNTDL 780
QY 774 LEDTPQCRICGGLAMVECRECYDDPDISAGKIQKQCTCTNTQVHLHPKRLNKHYNPVS 833
Db 781 LEDTPQCRICGGLAMVECRECYDDPDISAGKIQKQCTCTNTQVHLHPKRLNKHYNPVS 840
QY 834 PKDLPDWDNRHGGICPQNMELFAVLCTIETSHYVAFVKYKDDSAWLFFDSMAORDGGQNG 893
Db 841 PKDLPDWDNRHGGICPQNMELFAVLCTIETSHYVAFVKYKDDSAWLFFDSMAORDGGQNG 900
QY 894 FNIQVTPCPVEGYLKMSELDLHSLDSRRIOGCARRLLCDAYMCYQSTMSLYK 949
Db 901 FNIQVTPCPVEGYLKMSELDLHSLDSRRIOGCARRLLCDAYMCYQSTMSLYK 956

RESULT 3
Q8CGB0 PRELIMINARY; PRT; 952 AA.
AC Q8CGB0;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Similar to cylindromatosis (turban tumor syndrome).
OS 2010013M14RIK.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC042438; AAH42438.1; -.
DR MGD; MGI:1921506; 2010013M14RIK.
DR GO; GO:0004197; F.cysteine-type endopeptidase activity; IEA.
DR GO; GO:0004221; Fubiquitin thiolesterase activity; IEA.
DR GO; GO:0006511; Fubiquitin-dependent protein catabolism; IEA.
DR InterPro; IPR000938; Peptidase C19.
DR InterPro; IPR001394; CAP-Gly.
DR InterPro; IPR000408; Reg.chr_condens.
DR Pfam; PF01302; CAP_GLY; 2.
DR Pfam; PF00443; UCH; 1.
DR PROSITE; PS0245; CAP_GLY 2; 2.
DR PROSITE; PS00626; RCC1 2; 1.
DR PROSITE; PS0235; UCH 2.3; 1.
DR PROSITE; PS0235; UCH 2.3; 1.
SQ SEQUENCE 952 AA; 106585 MW; 0AC0C7D4FF215A9C CRC64;
```

Query Match 93.7%; Score 4718.5; DB 11; Length 952;
Best Local Similarity 94.2%; Pred. No. 0;
Matches 898; Conservative 22; Mismatches 28; Indels 5; Gaps 5;

```
QY 1 MSSGLMSEQKVTSPYWEERIFYLLQBCSVTDKQTKLLKVPKSGISGYIQDRSVGHRSI 60
Db 1 MSSGLMSEQKVTSPYWEERIFYLLQBCSVTDKQTKLLKVPKSGISGYIQDRSVGHRSI 60
QY 61 PSAGKKNQIGLTKLEQPHAVLFVDE-DVVEINEKFTELLAIITNCEERFSLFKNRRLS 119
Db 61 PSTKGNKNQIGLTKLEQPHAVLFVDEKDVVEINSEKFTELLAIITNCEERLSLFRNRLS 120
QY 120 KGLQIDVGCVPKQVLRSGEEKFPGVWRFRGPELLAERTVSGIFFGVLELEEGRGQGFQGV 179
Db 121 KGLQVDVGSVPKQVLRSGEEKFPGVWRFRGPELLAERTVSGIFFGVLELEEGRGQGFQGV 180
QY 180 YQKQLFQCDDECG-FVALDKLEIEDDDTALESIDYAGPDTMQVELPPLINSRVSLKG 238
Db 181 YQKQLFQCDDECGFVALDKLEIEDDDNGLESDFAGPDTMQVEPPPLEINSRVSLKV 240
QY 239 GETIESGTVIFCDVLPKESIGYFVGVDMPNPIGNWDGRFDGV-LCSFACVESTILLHN 297
Db 241 GESTESGTVIFCDVLPKESIGYFVGVDMPNPIGNWDGRFDGVQLCSFASVESTILLHN 300
QY 298 DIIPESVTQERRPPKLAFLMSRGVGDGSSSHNKPKATGSTSDPGR-RSELYFTLNGSSV 356
Db 301 DIIPDSVTQERRPPKLAFLMSRGVGDGSSSHNKPKVGTGSTSDPGRNRSELYFTLNGSSV 360
QY 357 DSQPSQKNTWYIDEVAEDPAKSLTEISTDFDRSSPPLOPPVNSLTFRHSLPFSL 416
Db 361 DSQ-QSKSNPWYIDEVAEDPAKSLTEMSDFGHSSPPPPQPPSMNLSSENRFHSLPFSL 419
QY 417 TKMPNTNGSGHSPSLSAQSVMEELNAPVQESPPPLAMPNGSHGLEVSLAEVKEP 476
Db 420 TKMPNTNGSHSPSLSAQSVMEELNAPVQESPPPLISSGNAHGLEVSLAEVKEP 479
QY 477 FYGVIRWIGOPPLGNEVLAGELEDEACAGCTDGTFRGTRVFTCALKKALFVKLKSQRDS 536
Db 480 FYGVIRWIGOPPLGNEVLAGELEDEACAGCTDGTFRGTRVFTCALKKALFVKLKSQRDS 539
QY 537 RFASLQPVSNQIERCNSLAFGGYLVSEVVEENTPPKMEKEGLEIMIGKKKGIOGHYNSCYL 596
Db 540 RFASLQPVSNQIERCNSLAFGGYLVSEVVEENTPPKMEKEGLEIMIGKKKGIOGHYNSCYL 599
QY 597 DSTLFCFLAFSSVLDTVLLRPKEKNDVEYVSETOELLARTEIVNPLRIYGYCATKIMKLR 656
Db 600 DSTLFCFLAFSSVLDTVLLRPKEKNDVEYVSETOELLARTEIVNPLRIYGYCATKIMKLR 659
QY 657 KILEKVEAASGFTSEEDKDPPEFLNLFPHILRVEPLKIRSAQKVDQCFYQIFMEKNE 716
Db 660 KILEKVEAASGFTSEEDKDPPEFLNLFPHILRVEPLKIRSAQKVDQCFYQIFMEKNE 719
QY 717 KVGVPITQQLLEWSFINSNLKFAEAPSLIIQMPRFQKPKLFPKIFPSLELNTDLLED 776
Db 720 KVGVPITQQLLEWSFINSNLKFAEAPSLIIQMPRFQKPKLFPKIFPSLELNTDLLED 779
QY 777 TPRQCRICGGLAMVECRECYDDPDISAGKIQKQCTCTNTQVHLHPKRLNKHYNPVSIPKD 836
Db 780 TPRQCRICGGLAMVECRECYDDPDISAGKIQKQCTCTNTQVHLHPKRLNKHYNPVSIPKD 839
QY 837 LPDMDNRHGGICPQNMELFAVLCTIETSHYVAFVKYKDDSAWLFFDSMAORDGGQNGFNI 896
Db 840 LPDMDNRHGGICPQNMELFAVLCTIETSHYVAFVKYKDDSAWLFFDSMAORDGGQNGFNI 899
QY 897 PQVTPCPVEGYLKMSELDLHSLDSRRIOGCARRLLCDAYMCYQSTMSLYK 949
Db 900 PQVTPCPVEGYLKMSELDLHSLDSRRIOGCARRLLCDAYMCYQSTMSLYK 952

RESULT 4
Q80TQ2 PRELIMINARY; PRT; 999 AA.
AC Q80TQ2;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE MKIAA0849 protein (Fragment).
```



```
Db 121 KGLQVDGSPVKYQLRSRGEKFGCVVFRGRLAERTVSGIFFGVELLEBGRGGFTDGV 180
Qy 180 YQKQLFQCDDECG-FVALDKLELIEDDDTALESYAGPGDTMQVELPPLPPEINSRVSLKG 238
Db 181 YQKQLFQCDDECGFVALDKLELIEDDDNGLESDFAGPGDTMQVEPPLPPEINSRVSLKV 240
Qy 239 GETIESGTWIFCDVLPKESISLGFVGVDMNDPIGNWMDGRFDGV-LCSFACVESTILLHIN 297
Db 241 GESTESGTWIFCDVLPKESISLGFVGVDMNDPIGNWMDGRFDGVLCFSFASVESTILLHIN 300
Qy 298 DIIP---ESVTQRRPPKLAFLMRGVDGKSSSHNKPKATGSTSDPCNR-RSELYFTLNG 353
Db 301 DIIPALSDSVTQRRPPKLAFLMRGVDGKSSSHNKPKVTCGSTDPGRNRSELYFTLNG 360
Qy 354 SSVDSQPSQSKNTWIDEVAEDPAKSLTEISDFDRSSPPLOPPVNSLTITENRFSHLP 413
Db 361 SSVDSQ-QSKSKNPWIDEVAEDPAKSLTEMSDDFGHSSPPQPSPVNSLSSENRFHSLP 419
Qy 414 FSLTKPNTNGSIGHSPSLSAQSVMEELNTPAVQESPPPLAMPNGSHGLEVGSLAEVKE 473
Db 420 FSLTKPNTNGSMAHSPSLSAQSVMEELNSTPVQESPPPLPISSGNAHGLEVGSLAEVKE 479
Qy 474 NPPFYGVIRWIGOPPGINEVLGLELEDEACAGCTDGTGTRFTYFTCALKKALFVKLXSCR 533
Db 480 NPPFYGVIRWIGOPPGINEVLGLELEDEACAGCTDGTGTRFTYFTCALKKALFVKLXSCR 539
Qy 534 PDSRFASLPVSNQIERCNSLAEGGYLSEVVEENTPPKMEKEGLEIMIGKKGIQGHYNS 593
Db 540 PDSRFASLPVSNQIERCNSLAEGGYLSEVVEENTPPKMEKEGLEIMIGKKGIQGHYNS 599
Qy 594 CYLDSTLFCFLAFSSVLDTVLLRP 617
Db 600 CYLDSTLFCFLAFSSALDTVLLRP 623
```

RESULT 6

```
Q9NZX9 ID Q9NZX9 PRELIMINARY; PRT; 515 AA.
AC Q9NZX9
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE HSPC057.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=20499367; PubMed=11042152;
RA Zhang Q.H., Ye M., Wu X.Y., Ren S.X., Zhao M., Zhao C.J., Pu G.,
RA Shen Y., Fan H.Y., Lu G., Zhong M., Xu X.R., Han Z.G., Zhang J.W.,
RA Tao J., Huang Q.H., Zhou J., Hu G.X., Gu J., Chen S.J., Chen Z.;
RT "Cloning and functional analysis of cDNAs with open reading frames for
RT stem/progenitor cells.";
RL Genome Res. 10:1546-1560 (2000).
DR EMBL; AF161542; AAF29029.1; -.
DR InterPro; IPR000938; CAP-Gly.
DR Pfam; PF01302; CAP_GLY; 1.
DR PROSITE; PS0245; CAP_GLY_2; 1.
SQ SEQUENCE 515 AA; 58197 MW; 23D82D7BD66EA146 CRC64;
```

Query Match 49.9%; Score 2512; DB 4; Length 515;
Best Local Similarity 94.3%; Pred. No. 2.6e-189;
Matches 477; Conservative 1; Mismatches 28; Indels 0; Gaps 0;

```
Qy 419 MNTNGSIGHSPSLSAQSVMEELNTPAVQESPPPLAMPNGSHGLEVGSLAEVKENPPFY 478
Db 1 MNTNGSIGHSPSLSAQSVMEELNTPAVQESPPPLAMPNGSHGLEVGSLAEVKENPPFY 60
Qy 479 GVIRWIGOPPGINEVLGLELEDEACAGCTDGTGTRFTYFTCALKKALFVKLXSCRPSRF 538
```

```
Db 61 GVIRWIGOPPGINEVLGLELEDEACAGCTDGTGTRFTYFTCALKKALFVKLXSCRPSRF 120
Qy 539 ASLPQVSNQIERCNSLAEGGYLSEVVEENTPPKMEKEGLEIMIGKKGIQGHYNSCYLDS 598
Db 121 ASLPQVSNQIERCNSLAEGGYLSEVVEENTPPKMEKEGLEIMIGKKGIQGHYNSCYLDS 180
Qy 599 TLFCLPAFSSVLDTVLLRPKEKNDVEYYSQELLRTIENVPLRIYGVYCATKIMKLKRI 658
Db 181 TLFCLPAFSSVLDTVLLRPKEKNDVEYYSQELLRTIENVPLRIYGVYCATKIMKLKRI 240
Qy 659 LEKVEAASGTSSEKDPPEFLNLFHLLRVEPPLKIRSAQKQVQDCYFYQIFMEKNEKV 718
Db 241 LEKVEAASGTSSEKDPPEFLNLFHLLRVEPPLKIRSAQKQVQDCYFYQIFMEKNEKV 300
Qy 719 GVPTIQQLLEWSFNSNLKFAEAPSCLIQMPRGKDFKPKIFPSLELNTDLELTP 778
Db 301 GVPTIQQLLEWSFNSNLKFAEAPSCLIQMPRGKDFKPKIFPSLELYIYILKTLTP 360
Qy 779 RQCRICGLAMEYECRECYDDPDISAGIKOPCKTCNTQVHLHPKRLNHNKYNPVSPLKDL 838
Db 361 DSAGYVGLQCMSVENATTIRTFSGAKIKOPCKTCNTQVHLHPKRLNHNKYNPVSPLKDL 420
Qy 839 DMDWRHGCIPQNNMELFAVLCIETSHYVAFVKYKDDSAWLFDFDSMADRDGGQNGFNIPQ 898
Db 421 DMDWRHGCIPQNNMELFAVLCIETSHYVAFVKYKDDSAWLFDFDSMADRDGGQNGFNIPQ 480
Qy 899 VTCPPEVGEYKMSLEDLHSLDSRRRI 924
Db 481 VTCPPEVGEYKMSLEDLHSLDSRRRI 506

RESULT 7
Q80VB3 ID Q80VB3 PRELIMINARY; PRT; 313 AA.
AC Q80VB3
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBDJ databases.
DR EMBL; BC049879; AAH49879.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 313 AA; 36198 MW; BED5662EAA7453DA CRC64;
```

Query Match 33.0%; Score 1662; DB 11; Length 313;
Best Local Similarity 97.8%; Pred. No. 1.4e-122;
Matches 306; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```
Qy 637 IVNPLRIYGVYCATKIMKLKILEKVEAASGFTSEKDPPEFLNLFHLLRVEPPLKIR 696
Db 1 IVNPLRIYGVYCATKIMKLKILEKVEAASGFTSEKDPPEFLNLFHLLRVEPPLKIR 60
Qy 697 SAQKQVQDCYFYQIFMEKNEKVGPVTIQQLLEWSFNSNLKFAEAPSCLIQMPRGKDF 756
Db 61 SAQKQVQDCYFYQIFMEKNEKVGPVTIQQLLEWSFNSNLKFAEAPSCLIQMPRGKDF 120
Qy 757 KLFKKIFPSLELNTDLETPQCRICGGLAMEYECRECYDDPDISAGIKQFCKTQNTQ 816
Db 121 KLFKKIFPSLELNTDLETPQCRICGGLAMEYECRECYDDPDISAGIKQFCKTQNTQ 180
Qy 817 VHLHPKRLNHNKYNPVSPLKDLPDMDWRHGCIPQNNMELFAVLCIETSHYVAFVKYKDD 876
Db 181 VHLHPKRLNHNKYNPVSPLKDLPDMDWRHGCIPQNNMELFAVLCIETSHYVAFVKYKDD 240
```


RA	Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,	
RA	Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,	
RA	Asburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;	
RT	"Annotation of Drosophila melanogaster genome.";	
RL	Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.	
RN	[4]	
RP	SEQUENCE FROM N.A.	
RA	Adams M.D., Celnikier S.E., Gibbs R.A., Rubin G.M., Venter C.J.;	
RL	Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.	
RS	[5]	
RN	SEQUENCE FROM N.A.	
RA	FlyBase;	
RL	Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.	
DR	EMBL; AEO03628; AAM10741.1; -.	
DR	FlyBase; FBGN0032210; CG5603.	
DR	GO: GO:0004197; F:cysteine-type endopeptidase activity; IEA.	
DR	GO: GO:0004321; F:ubiquitin thioesterase activity; IEA.	
DR	GO: GO:0006151; P:ubiquitin-dependent protein catabolism; IEA.	
DR	InterPro: IPR001394; Peptide_C19.	
DR	Pfam: PF00443; UCH; 1.	
DQ	SEQUENCE 551 AA; 62315 MW; 9D4DG3B287B7783 CRC64;	
Query Match 22.6%; Score 1136.5; DB 5; Length 551;		
Best Local Similarity 45.08; Pred. No. 9.8e-81;		
Matches 244; Conservative 90; Mismatches 175; Indels 33; Gaps 15;		
QY	430 PLSL-----SAQSVMEEELNTAPVSPESPLAMPNGSHGLEVSIAEVKENP----PFYGVI 481	
DB	: : : : : :	
DB	19 PLTVLEFYSPLLPEELGFT-IADYQLPIEIPGT-E--LAIGSLVEV-SNPGVCEDLYGVV 74	
QY	482 RWTGQQPG-LNEVLAGLEDEEC-----AGCTDGTFRGTRYPTCAKHALFKLKSCRPS 536	
DB	: : : : : :	
DB	75 RWGIIPPQGOKNLVGIEVEDSELNKNVASDGRHNGVRFLTCHDGRAIFVPANRCTADR 134	
QY	537 RFASLQP--VSNOIERCNSLAFGCYLVSEVVENTPPKM--EKEGLETIMGKKKGIOGHYN 592	
DB	: : : : : :	
DB	135 RFADVNDNISANRVSSNHAKKFGVADCPAYGISLPPIQIHNSDDELASICGKFKGIOGHNN 194	
QY	593 SCYLDSTFLCLFAFSNVLDITVLLRKPKEKDVEYYSETQELLRTIVNPLRIYGYVCATKI 652	
DB	: : : : : :	
DB	195 SCYLDATLFMSFTTSVFDSILYRRPGPDIRNYSEVKVLRDENVNPLRNWVRSDRV 254	
QY	653 MKLRILEKYEAASGFTSBEKBDEEFENLIUHFHLIRVEPLLKIRSAGKVQCDCYFYQIFM 712	
DB	: : : : : :	
DB	255 MKLRELIDQLSSVSGLTCEEKDPDEFINSLLSQIMRVEPEPKLSGG---QDSYFYQLFV 310	
QY	713 EKKEKVGVTIQOLLEWSFNISNLKFAEPASCIIOMPREGDKPLFKKIFPSLEINIID 772	
DB	: : : : : :	
DB	311 EKDEKLTPSVQQLFEOSFHSSDIKLKEVPSCFI IQMPRFCKNYKMYPRILFPQGLVDTD 370	
QY	773 LLEDTPQRICRGGLAMYCRECVDDPD1SAG-KIKOFCKTCTQVHLHKPLNHHKNPV 831	
DB	: : : : : :	
DB	371 IIENSFRCSLCGLAEYECRCDFGSLOAGSLECTAFCPKCLKTFASHIKRTNHSVSKI 430	
QY	832 SLRKDLFDWRHGCI PCQNMEELPAVLICETISHYVAFFVKYGK-DDSAWLFFDSMDRGG 890	
DB	: : : : : :	
DB	431 YSPKEFKIMA-EHMVVPRLMYELFAVVICETISHYVAFVKSGSGDAPWCFFDSMDRKE 489	
QY	891 QNGFNIPQVTPCEVGEYLK-----MSLEDLHSLSRSRIQCARRLLCDAYNCMVQSTMS 946	
DB	: : : : : :	
DB	490 QNGYNIPEITCVPELTOWLSEEARSINET-STNDKVLPHAKRIFCDAYMCLEYQSTDIM 548	
QY	947 LY 948	
DB	:	
DB	549 MY 550	
RESULT 10		
ID	G8IPC5 PRELIMINARY; PRT; 639 AA.	
AC	QBPC5;	
DT	01-MAR-2003 (TrEMBLrel. 23, Created)	
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)	


```
QY 208 TALESDYAGPDTMQWELPPLLEI-----NSRVS--LKGG-----ET 241
DQ : : : : :
Db 129 IRFSSNYGREBPCVIEIPGTVREMAADDWKMSELKEWFTKSRASSHLRDGLAMPWEE 188
QY 242 IESGTWIFC-----DVLPGKESLGVFGVDMNPICNWDGDFDGLVCSFACVESTILL 294
DQ : : : : :
Db 189 LDC-TPLICAMITRSDVMRINQDAIHLAVSEKRI-----EYQNFENFNFILNL 239
QY 295 HINDIIPESV--TQERPPPKLAFMRSGVDKG-----SSHNKP 331
DQ : : : : :
Db 240 KIGDSVSEVEDTMRVPKAVKSWIGERPEASGTYWNVDPGNTSQWPSSNQSSSHDRL 299
QY 332 -----KATGSTS-DPGNRRSELYTLN-----GSSVDS 358
DQ : : : : :
Db 300 NRQFDTNWNPFGSSSVAPSN--SRLYSPNQMHMPKGGVSALYDNRRLVQVSGDEE 357
QY 359 -----OPOSKSNWYIDEVAEDPAKSLTEISTDPR--SSPLQP 397
DQ : : : : :
Db 358 QYRSAPKAPRERIIIPVSRQOQPIEQRN-----SRSMKPSPDYNTYSTHPPRP 406
QY 398 PPVNSLT--TENRPHSLPFLSTKMPNT--NGSIGHSPLSLSAQSVMEELNAPVOESP 452
DQ : : : : :
Db 407 PPSMMNYPSMWSNTHSLQPSRSKVQTIQRNQVRQNPARNRVEDSDQNFRIQDQC-- 464
QY 453 LAMPNGSHGLEVGSALAEVKENPPFYGVIRWIGQPPGLNEVLAGLELEDECAGCTDGR 512
DQ : : : : :
Db 465 ----IWNNGAEB-----RGLIKYIGLKGKTLYAGVFPKN--TIGAGTGVPN 507
QY 513 GTRYFTCALKKALFVKLKCRC--PDRPASPLOPVSNQIERCNLSLAFGGYLSEVVEENTPPK 571
DQ : : : : :
Db 508 REQFLAKDGHAGFVLSLSPSSISSTSTSSQHR-----RLSSRSQQMP-- 557
QY 572 MEKEGLEIMTKGKKGIOGHVNSCYLSTLFLC-----FAPSVLDTVLLRPEKNDVE 624
DQ : : : : :
Db 558 -AASGTSISV-----PVNGRHNVNGKQASIESLSDPPPPYAPPSPRASLLEKSIGS-E 611
QY 625 YVSETQELARTEIVNPLRIYGYCATKIMKRLIL--EKVEAASGFTSEKDPPEFLNLF 683
DQ : : : : :
Db 612 TAQOFKILAEIVFLPKRVHYVRAHVMKRLKLAELMHPVTLGTLNEEKDPEILGFIF 671
QY 684 HHILRVEPLLKRSAGOKVQDCYFYQIFMEKNEKVGVTPTIQQLLEWSFINSNLKFAEAPS 743
DQ : : : : :
Db 672 SKVFHAEPFKLTIGONHAKDSQYLPVVD--DWLGAATSOHLERHMRSAQVTFKAPP 730
QY 744 CLIIQMPREKDFKPKFKLPPSLEMLITDLETPQCRCIGGLAMEYECRECYDDDISA 803
DQ : : : : :
Db 731 VLIMQLPRYGOQ--KVFDKILPLETIDITPFVAGAVPACSKQACSEVYCTCTLTRRFF 789
QY 804 GKIKQCKTCNTQVHLHPKRLNHYN--PVSLPKDLPDWDWRHGCIPQNMELFVLICI 860
DQ : : : : :
Db 790 SEV-TECRKCFHTHLLPEIEDHKSRLDLYPPGPKQKPH-----SHKWLSAVLCI 839
QY 861 ETSHYVAFVYKGDSDAWLFFDSMADRDGQONGFNIPOVTPCEVGEYLAWS-----LEDL 916
DQ : : : : :
Db 840 ETSHYVAYVR--TSSNQWFFDSMADREGLSDGFNVVRECGNMSDWLSLQGNRLKOA 897
QY 917 H-----SLDSARIO-----GCARLLCDAYCMWQ--SPTWS 946
DQ : : : : :
Db 898 DECGQIKWELNKNISPNDRKLEIAMFGKOSSLDPLVGRLLSDSYCYFVEDASPTSS 954
```

RESULT 15

OSS156
ID AC PRELIMINARY; PRT; 1046 AA.

DT 01-JUN-1998 (TEMBLrel. 06, Created)
DT 01-JUN-1998 (TEMBLrel. 06, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
DE CLIP-115 protein.
GN CLIP-115.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98087115; PubMed=9427243;
RA de Zeeuw C.I., Hoogenraad C.C., Goedknecht E., Hertzberg E.,
RA Neubauer A., Grosveld P.G., Galjart N.J.;
RT "CLIP-115, a novel brain specific cytoplasmic linker protein, mediates
RT the localisation of dendritic lamellar bodies.";
RL Neuron 19:1187-1199(1997).
DR EMBL; AJ000485; CAA04123.1; -
DR PIR; T42734; T42734.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0005667; C:transcription factor complex; IEA.
DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. .; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0006367; P:transcription initiation from Pol II promoter; IEA.
DR InterPro; IPR000938; CAP-Gly.
DR InterPro; IPR000812; TFIIB euk.
DR Pfam; PF01302; CAP GLY; 2.
DR PROSITE; PS00845; CAP GLY 1; 2.
DR PROSITE; PS00245; CAP GLY 2; 2.
DR PROSITE; PS00782; TFIIB; 1.
SQ SEQUENCE 1046 AA; 115480 MW; 72B6CE9F76D2A1D0 CRC64;

Query Match 3.0%; Score 149.5; DB 11; Length 1046;
Best Local Similarity 20.9%; Pred. No. 0.021;
Matches 115; Conservative 66; Mismatches 189; Indels 179; Gaps 24;

QY 309 RPKLAPMSRGVGDGSSSHNPKATGSTSDPGNRRSELYTLNGSSVDSQ-----PQ 361
DQ : : : : :
Db 8 KPP-----GRGKHSVPGRPSI-----GSASSVVASASGSGSLPHKQASGPS 53
QY 362 SKSNWYIDEVAEDPAKSLTEISTDP-----DRSPPLQPPVNSLTENRF----- 409
DQ : : : : :
Db 54 SAGATT-----TVSEKPGPKAAEVGDGDFGVGVGVNGVKGQVQVLTGETQAPQGW 109
QY 410 -----HSLPFSLTQMP--NTNGSIGHSPLSLSAQ 437
DQ : : : : :
Db 110 AGVVDLPVGNKDNAGVGLRYFPCPALQGIIFTRPSKLTQRPAAGSGSDGSHVESLTAQN 169
QY 438 VMEELNAPVOESPPL-----AMPNGSHGLEVGSALAEVKENPPFY----- 478
DQ : : : : :
Db 170 LSLHSGTA-----TPPLTGRVPLRESVLNNSVKTGNSGSLSDSGSVKRGDKDLHLGDR 225
QY 479 -----GVIRWIGQPPGLNEVLAGLELEDECAGCTDGTFRGTRYTCALKKALFVKL- 529
DQ : : : : :
Db 226 VLVGTKTGVVRYVGETDFAKGECWGVEL--DEPLGKNDGAVAGTRYFCQCPKFLGFIADIH 284
QY 530 -----KSCRPSRFA-----SLQPVSNQIERCNLSLA--FGGYLSEVVEENTP 569
DQ : : : : :
Db 285 KVRIGFSPSTSPAKAKTKRMAMGVSAHSPSSSSISSVSSVASSVGGRRPS----- 336
QY 570 PRMEKEGLEIMTKGKKGIOGHVNSCYLSTLFLCFAFSSVLDTVLLRP--KEKND-VEYY 626
DQ : : : : :
Db 337 -----RSGLLTETSSRYA-----RKISGTTALQELAKEQKHIEQL 372
QY 627 SETQELLRLTEIVNPLRIYGVCAKIMKRLKLEKVEAA--SGFTSEKDPPEFLNLFPHH 685
DQ : : : : :
Db 373 LAERDLERAERVAK-----ATSHICEVEKEIALLKAQHEQYVAEAEKQLQARLLVEN 424
QY 686 IL--RVEPLLKRSAGOKVQDCYFYQIFMEKNEKVGVTPTIQQL-----LEWSFINSNL 736
DQ : : : : :
Db 425 VRKEKVDLSNQLSEERKVEDLQF--RVEEESITKGDLETQQLSHARIGELESQLL---L 480
QY 737 KFAEAPSCL 745
DQ : : : : :
Db 481 EKAQAEELL 489

Search completed: September 17, 2004, 09:48:16
Job time : 148 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 18, 2004, 15:16:48 ; Search time 8073 Seconds
(without alignments)
5095.075 Million cell updates/sec

Title: US-09-671-687A-3
Perfect score: 5034
Sequence: 1 MSGSLWSQKVTSPYWEERI.....RLLCDAYMCYQSPMTSLYK 949

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -MODEL=frame_p2n.model -DRV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09671687/runat_17092004_103034_2781/app.query.fasta_1.1095
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09671687@cgn_1_15932@runat_17092004_103034_2781 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCKS=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4983	99.0	3540	9	BC012342 Homo sapi
2	4983	99.0	5414	9	AB020656 Homo sapi
3	4971.5	98.8	5371	9	HSA250014 Homo sapi
4	4942	98.2	4527	6	BD231207 Human cyt
5	4718.5	93.7	4314	10	AK122389 Mus muscu
6	4718.5	93.7	4501	10	BC042438 Mus muscu
7	4601	91.4	2845	6	AX883937 Sequence
8	4601	91.4	2845	6	BD160617 Primer fo
9	4601	91.4	2845	6	AK024348 Homo sapi
10	4600.5	91.4	241990	2	AC123449 Rattus no
11	4600.5	91.4	241990	2	AC098162 Rattus no
12	3676	73.0	2523	6	AR338799 Sequence
13	3433.5	68.2	2341	9	AK000187 Homo sapi
14	3022.5	60.0	2569	6	AX883681 Sequence
15	3022.5	60.0	2569	6	BD160470 Primer fo
16	3022.5	60.0	2569	9	AK024212 Homo sapi
17	2930	58.2	1954	9	AK056226 Homo sapi
18	2493	49.5	241630	2	AC131881 Rattus no
19	2468	49.0	2116	6	AK017216 Sequence
20	2468	49.0	2116	6	BD135433 Receptor
21	2108.5	41.9	215717	2	AC133651 Mus muscu
22	1736	34.5	251957	2	AC126867 Rattus no
23	1662	33.0	2426	10	BC049879 Mus muscu
24	1442.5	28.7	212531	2	AC145178 Gallus ga
25	1430.5	28.4	129025	2	AC145180 Gallus ga
26	1262	25.1	837	6	AX869679 Sequence
27	1262	25.1	837	6	BD149741 Primer fo
28	1215	24.1	758	6	AX867014 Sequence
29	1215	24.1	758	6	BD147076 Primer fo
30	1132.5	22.5	2068	3	AY071592 Drosophil
31	1111	22.1	151567	2	AC145503 Canis fam
32	1111	22.1	199277	2	AC145445 Canis fam
33	1109.5	22.0	208061	2	AC145254 Bos tauru
34	1108.5	22.0	168271	9	AC007728 Homo sapi
35	1097.5	21.8	141663	2	AC145018 Felis cat
36	1095.5	21.8	163319	9	HSA303140 Homo sapi
37	1085	21.6	155857	2	AC145174 Pan trogl
38	1085	21.6	178790	2	AC145238 Pan trogl
39	1078	21.4	194869	2	AC145003 Papio anu
40	1077.5	21.4	188742	2	AC138025 Mus muscu
41	1067	21.2	232509	2	AC134093 Rattus no
42	936	18.6	84367	3	AC005454 Drosophil
43	936	18.6	135011	2	AC020183 Drosophil
44	936	18.6	153505	3	AC009849 Drosophil
45	936	18.6	154779	3	AC092237 Drosophil

ALIGNMENTS

BC012342	BC012342	3540 bp	mRNA	linear	PRI 04-OCT-2003
LOCUS	Homo sapiens cylindromatosis (turban tumor syndrome), mRNA (cDNA				
DEFINITION	clone MGC:19923 IMAGE:4552767), complete cds.				
ACCESSION	BC012342				
VERSION	BC012342.1	GI:15214433			
KEYWORDS	MGC.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 3540) Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altshul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, I., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Udén, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Lequellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Halyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, A.S., Krzyzanski, M.I., Skalska, U., Smalhus, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.				
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)				
MEDLINE	22388257				
PUBMED	12477932				
REFERENCE	2 (bases 1 to 3540) Straussberg, R.				
AUTHORS	Direct Submission				
TITLE	Submitted (15-AUG-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA				
JOURNAL	NIH-MGC Project URL: http://mgc.nci.nih.gov				
REMARK	Contact: MGC help desk				
COMMENT	Email: cgabs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada info@bcgsc.bc.ca Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzyzanski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Fawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schain, Duane Smalley, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.				
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAL Plate: 29 Row: c Column: 5 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 14165257. Location/Qualifiers 1..3540 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="MGC:19923 IMAGE:4552767" /tissue_type="Uterus, leiomyosarcoma" /clone_lib="NIH_MGC_46"				
source					
gene	/lab_host="DH10B-R" /note="Vector: pOTB7" 1..3540 /gene="CYLD" /db_synonyms="CDMT, CYLD1, EAC, KIAA0849, HSPC057" /note="synonyms: CDMT, CYLD1, EAC, KIAA0849, HSPC057" /db_xref="LocusID:1540" /db_xref="MIM:605018" 277..3138 /codon_start=1 /product="CYLD protein" /protein_id="AAH12342.1" /db_xref="GI:15214434" /db_xref="LocusID:1540" /translation="MSSSLGWSQEKVTSYWEERIFVLLLOEQSVTDKOTQKLLKVPKG SIGYIQDRSVGHSRIRPSAKGKKNGIKLLEQPHAVLFVDEKDVVEINEKTELLLA ITNCERSLFKNNRRLKGLIDVGVKQVLRGEEKFPVGRFGLPALTIVSG IPFGVLEEGRGQFTDVGQKQLFCQDEDCGVFVADLDKLELIDDDTALAESDYAG PDMTQVLELPLEINRSVLKGETIESGTIVFCVLPKESLGIFVGVDMNPIGNW KDRFDVQLCSFACVESVILLINDIIIPESVTOERRPKLAFMSRGVKGKSSHNKP KATGSTDPGNRNSRSELFVTLNGSSVDSPQSKNTWYIDEVADPPAKSLTEISDF DRSSPLPPPPVNSLTTRRHSFLPSLTMTNGTSGHSPLSLSAQSVMBELNTP VQESPLAMPNGSHGLEVSLAEVKENPPFYGVIRWIGQPPGLNEVLAGELEDECA GCTDGTFRGTFRFTALKKALFVLKLSRCPDRSFASLQVSNQIERCNSLAFGGYLSE VVEENTPPMEKEGLEIMIGKKGIQGHYNSCYLSDTLFCFLFASSVLDTVLLRPEK NDVEYSETQLLEIIVNPLRIYGVVCAIKMLKLEKILKLEFAASGFTSEKDPDEF LNILPHILRVPELLKIPRSGQVQDCYFOIEMEKNEKVGVTTCQLLEWFSINSL KFAEASCLIIQMPRFGDKFLFKIPFSLNITDLEDTPRQRCGLGAYECRE CYDDPDISAGKIKQCKTCTQVHLHPKLNHNKYNPVSPLKDLDPMDMRHGCI PQNM ELFVLCIETSHYVAFVRYGKDSAWLFFDSMADRGQGNFNIQVTPCPPEVGEYLK MSLEDLSLSRRIQGCARRLLCDAYMCWYQSPMTSLVK" 745..867 /note="CAP_GLY; Region: CAP-Gly domain. CAP stands for cytoskeleton-associated proteins" /db_xref="CDD:pfam01302" 1720..2025 /note="NIP100; Region: COG5244, NIP100, Dynactin complex subunit involved in mitotic spindle partitioning in anaphase B [Cell division and chromosome partitioning]" /db_xref="CDD:COG5244"				
misc_feature	Alignment Scores: Pred. No.: 0 Length: 3540 Score: 4983.00 Matches: 948 Percent Similarity: 99.48% Conservative: 0 Best Local Similarity: 99.48% Mismatches: 1 Query Match: 98.99% Indels: 4 DB: 9 Gaps: 4 US-09-671-687A-3 (1-949) x BC012342 (1-3540) QY 1 MetSerSerGlyLeuTrpSerGlnGluysValThrSerProTyrTrpGluArgile 20 Db 277 ATGAGTTTCAGGCTTATGGAGCCCAAGAAAAGTCACTTCACTTGGGAAGCGGATT 336 QY 21 PheTyrLeuLeuGlnGluCysSerValThrAspLysGlnThrGlnLysLeuLys 40 Db 337 TTTTACTTCTTCTTCAAGATGCGCGTTACAGACAAACAAACAAAGCTCTTAA 396 QY 41 ValProLysGlySerLeuGlyGlnTyrLeuGlnAspArgSerValGlyHisSerArgile 60 Db 397 GTACCGAAGGGAAGTATAGACAGATATATCAAGATCGTTCGTGGGGCATTCAGGATT 456 QY 61 ProSerAlaLysGlyLysLysAsnGlnLeuGlyLeuLysLeuGlnProHisAla 80 Db 457 CCTTCTGCAAAAGGCAAGAAAATCAGATTGATTAATAATTTCTAGAGCAACCTCATGCA 516 QY 81 ValLeuPheValAspGlu---AspValValGluLeuLeuAsnGluLysPheThrGluLeuLeu 99 Db 517 GTTCTCTTTTGTGATCAAAAGGATGTTGTAGAGATAAATAAGTTCACAGATTACTT 576 QY 100 LeuAlaLeuThrAsnGluGluGluArgPheSerLeuPheLysAsnArgAnArgLeuSer 119				

Db 577 TTGGCAATTACCAATTGTGGAGGAGGTTGAGCCTGTTTAAAAACAGAAACAGACTAAGT 636
Qy 120 LysGlyLeuGlnIleAspValGlyCysProValLysValGlnLeuArgSerGlyGluGlu 139
Db 637 AAAGGCTCCAAATAGACGTGGGCTGCTCTGTGAAGTACAGCTGAGATCTGGGGAAGAA 696
Qy 140 LysPheProGlyValValArgPheArgGlyProLeuLeuAlaGluArgThrValSerGly 159
Db 697 AAATTTCTCTGGAGTTGTACGCTTCAGAGGACCCCTGTTAGCAGAGACAGTCTCCGGA 756
Qy 160 IlePhePheGlyValGluLeuLeuGluGluGlyArgGlyGlnGlyPheThrAspGlyVal 179
Db 757 ATATTTCTTGGAGTTGAAATGCTGGAAGAAGGTCGTGTCGAAGTTTCACTGACGGGGTG 816
Qy 180 TyrGlnGlyLysGlnLeuPheGlnCysAspGluAspCysGly---PheValAlaLeuAsp 198
Db 817 TACCAAGGAAACAGCTTTTTCAGTGTGATGAAGATTGTGGCGTGTGTTGTCATTGGAC 876
Qy 199 LysLeuGluLeuIleGluAspAspThrAlaLeuGluSerAspTyrAlaGlyProGly 218
Db 877 AAGCTAGAACTCATAGAAGATGATGACACTGCATTGGAAAAGTGATTACGCGAGTCTCTGGG 936
Qy 219 AspThrMetGlnValGluLeuProProLeuGluIleAsnSerArgValSerLeuLysGly 238
Db 937 GACACAATGCAGGTCGAACCTTCCTCTTGGAAATAAACTCCAGAGTTTCTTTGAAGGTT 996
Qy 239 GlyGluThrIleGluSerGlyThrValIlePheCysAspValLeuProGlyLysGluSer 258
Db 997 GGAGAAACAATAGAACTGGAACAGTTATATTCGTGATGTTTGGCAGGAAAGAAAGC 1056
Qy 259 LeuGlyTyrPheValGlyValAspMetAspAsnProIleGlyAsnTrpAspGlyArgPhe 278
Db 1057 TTAGGATATTTGTGTGTGGACATGGATAACCTATTGGCAACTGGGATGGAAGATT 1116
Qy 279 AspGlyVal---LeuCysSerPheAlaCysValGluSerThrIleLeuLeuHisIleAsn 297
Db 1117 GATGGAGTGCAGCTTTGTAGTTTTCGTGTGTTGAAAGTACAATTCTATTGCACATCAAT 1176
Qy 298 AspIleLeuProGluSerValThrGlnGluArgProProLysLeuAlaPheMetSer 317
Db 1177 GATATATCCAGAGGTGACGAGAAAGGAGGCTCCCAAACTTGGCTTTATGTCA 1236
Qy 318 ArgGlyValGlyAspLysGlySerSerHisAsnLysProLysAlaThrGlySerThr 337
Db 1237 AGAGGTGTGGGACAAAGGTTTCATCCAGTCATAATAAACCAAGAGGTACAGGATCACC 1296
Qy 338 SerAspProGlyAsnArg---ArgSerGluLeuPheTyrThrIleAsnGlySerSerVal 356
Db 1297 TCAGACCTCGAAATAGAAACAGATCTGAATTTATTTATACCTTAAATGGGTCTCTGTT 1356
Qy 357 AspSerGlnProGlnSerLysSerLysAsnThrTrpTyrIleAspGluValAlaGluAsp 376
Db 1357 GACTCACAAACCAATCCAATCAAAATAATACATGGTACATTGATGAAGTTGCAGAAAGAC 1416
Qy 377 ProAlaLysSerLeuThrGluIleSerThrAspPheAspArgSerSerProProLeuGln 396
Db 1417 CCTGCAAAATCTCTTACAGAGATATCTACAGACTTTGACCGTCTTCACCAACCACTCCAG 1476
Qy 397 ProProProValAsnSerLeuThrThrGluAsnArgPheHisSerLeuProPheSerLeu 416
Db 1477 CCTCCTCTGTGAACCTCACTGACCACGAGAACAGATTCCACTCTTTTACCATTCACTCTC 1536
Qy 417 ThrLysMetProAsnThrAsnGlySerIleGlyHisSerProLeuSerLeuSerAlaGln 436
Db 1537 ACCAAGATGCCCAATCAATGAAGTATTGGCCACAGTCCACTTCTCTGTGACGCCAG 1596
Qy 437 SerValMetGluLeuLeuAsnThrAlaProValGlnGluSerProProLeuAlaMetPro 456
Db 1597 TCTGTAATGGAAGAGCTTAAACACTGCACCCGTCAGAGAGTCCACCTTGGCCATGCCT 1656
Qy 457 ProGlyAsnSerHisGlyLeuGluValGlySerLeuAlaGluValLysGluAsnProPro 476
Db 1657 CTGGGAACTCACATGGTCTAGAGTGGGCTCATTTGGCTGAAGTTAAGGAGAACCCCTCT 1716

Qy 477 PheTyrGlyValIleAArgTrpIleGlyGlnProProGlyLeuAsnGluValLeuAlaGly 496
Db 1717 TTCTATGGGTAAATCCGTTGGATCGGTGATCCAGCCAGGACTGAATGAAGTGTCTCGTGA 1776
Qy 497 LeuGluLeuGluAspGluCysAlaGlyCysThrAspGlyThrPheArgGlyThrArgTyr 516
Db 1777 CTGGAACCTGGAAGATGAGTGTGACGGCTGTACGGATGGAACCTTCAGAGGACACTCGGTAT 1836
Qy 517 PheThrCysAlaLeuLeuLysLysAlaLeuPheValLysLeuLysSerCysArgProAspSer 536
Db 1837 TTCACTGTGCGCTGAAAGAGCGCTGTTGTGAAACTGAAAGAGCTGAGGCTGACACTCT 1896
Qy 537 ArgPheAlaSerLeuGlnProValSerAsnGlnIleGluArgCysAsnSerLeuAlaPhe 556
Db 1897 AGGTTTCATCATTTGCAGCGGTTTCCAAATCAGATTGAGCGCTGTAACCTCTTTAGCAATT 1956
Qy 557 GlyGlyTyrLeuSerGluValValGluGluAsnThrProProLysMetGluLysGluGly 576
Db 1957 GGAGGCTACTTAAGTGAAGTAGTAGAAATAATACTCCACCAAAATGGAAGAAAGGC 2016
Qy 577 LeuGluIleMetIleGlyLysLysGlyIleGlnGlyHisTyrAsnSerCysTyrLeu 596
Db 2017 TTGGAGATAATGATTGGGAAGAAAGAGCATCCAGGGTCAATTACAAATCTTGTGTACTTAA 2076
Qy 597 AspSerThrLeuPheCysLeuPheAlaPheSerSerValLeuAspThrValLeuLeuArg 616
Db 2077 GACTCAACCTTATTTCTGCTTATTTGCTTTTAGTTCTGTCTGGACACTGTGTACTTAA 2136
Qy 617 ProLysGluLysAsnAspValGluTyrTyrSerGluThrGlnGluLeuLeuArgThrGlu 636
Db 2137 CCCAAAGAAAGAACGATAGTAATAATATAGTGAACCCCAAGAGCTACTGAGGACAGAA 2196
Qy 637 IleValAsnProLeuArgIleTyrGlyTyrValCysAlaThrLysIleMetLysLeuArg 656
Db 2197 ATTGTTAATCCTCGAGATATATGGATATGTGTGTGTCACAAAAATTTATGAACCTGAGG 2256
Qy 657 LysIleLeuGluLysValGluAlaAspGlyPheThrSerGluGluLysAspProGlu 676
Db 2257 AAATACTTGAAGAGTGGAGGCTGCATCAGATTACCTCTGAAGAAAAAGATCCTGAG 2316
Qy 677 GluPheLeuAsnIleLeuPheHisIleLeuArgValGluProLeuLeuLysIleArg 696
Db 2317 GAATTTCTGAATATTCGTTTTCATCATATTTAAAGGTAGAACCTTTCTCTAAAAATAAGA 2376
Qy 697 SerAlaGlyGlnLysValGlnAspCysTyrPheTyrGlnIlePheMetGluLysAsnGlu 716
Db 2377 TCAGCAGGTCAAAAGGTACAAGATTGTTACTTCTTATCAAAATTTTATGGAAGAAAAATGAG 2436
Qy 717 LysValGlyValProThrIleGlnLeuLeuGluTrpSerPheIleAsnSerAsnLeu 736
Db 2437 AAAGTTGGCGTTCACCAATTCAGCAGTGTGTTAAGATGGTCTTTTATCAACAGTAACCTG 2496
Qy 737 LysPheAlaGluAlaProSerCysLeuIleIleGlnMetProArgPheGlyLysAspPhe 756
Db 2497 AAATTTGCAGAGGACCATCATGTCTGATTATTTCAGATGCCTCGATTTTGGAAAGACTTT 2556
Qy 757 LysLeuPheLysIlePheProSerLeuGluLeuAsnIleThrAspLeuLeuGluAsp 776
Db 2557 AAACATTTAAAAAAATTTTCTCTCTCTGGAATTAATAATAACAGATTTTACTTGAAGAC 2616
Qy 777 ThrProArgGlnCysArgIleCysGlyGlyLeuAlaMetTyrGluCysArgGluCysTyr 796
Db 2617 ACTCCCAAGACGTGCCGATATGAGGAGGCTTCCAAATGTATGATGTAGTAGAATGCTAC 2676
Qy 797 AspAspProAspIleSerAlaGlyLysIleLysGlnPheCysLysThrCysAsnThrGln 816
Db 2677 GACATCCGACATCTCAGCTGGAAAAATCAAGCAGTTTTTGTAAACCTGCAACTCAA 2736
Qy 817 ValHisLeuHisProLysArgLeuAsnHisLysTyrAsnProValSerLeuProLysAsp 836
Db 2737 GTCCACCTTCATCCGAGAGGCTGAATCATATAATAAACCCAGTGTCTACTTCCCAAGAC 2796

QY 837 LeuProAspTTPAspTTPArgHisGlyCysIleProCysGlnAsnMetGluLeuPheAla 856
 Db 2797 TTAACCGACTGGAGTGGAGACACGGCTGCATCCCTTCCAGAAATATGAGTTATTTGCT 2856
 QY 857 ValLeuCysIleGluThrSerHisTyrValAlaPheValLysTyrGlyLysAspSer 876
 Db 2857 GTTCTCTGATAGAAACAGACCACTATGTTGCTTTTGTGAAGTATGGAAGACCATTTCT 2916
 QY 877 AlaTTPLeuPhePheAspSerMetAlaAspArgAspGlyGlyGlnAsnGlyPheAla 896
 Db 2917 GCCTGGCTCTTTTACAGACATGCGGATGGGATGGTGGTCAAGTGGCTTCAACATT 2976
 QY 897 ProGlnValThrProCysProGluValGlyGluTyrLeuLysMetSerLeuGluAspLeu 916
 Db 2977 CTTCAAGTCAACCCATGCCAGAGTAGAGAGTACTTGAAGATGCTCTGGAAGACCTG 3036
 QY 917 HisSerLeuAspSerArgArgIleGlnGlyCysAlaArgArgLeuLeuCysAspAlaTyr 936
 Db 3037 CATTCCTTGAGCTCCAGAGAGATCCAAAGCTGTGCAGGAAGACTGCTTTGTGATGCATAT 3096
 QY 937 MetCysMetTyrGlnSerProThrMetSerLeuTyrLys 949
 Db 3097 ATGTGCATGTACAGAGTCCAAATGAGTTTGTACAAA 3135
 RESULT-2
 Locus AB020656 5414 bp mRNA linear PRI 17-MAY-2001
 DEFINITION Homo sapiens mRNA for KIAA0849 protein, partial cds.
 ACCESSION AB020656
 VERSION AB020656.2 GI:14133220
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (sites)
 Nagase, T., Ishikawa, K., Suyama, M., Kikuno, R., Hirose, M.,
 Miyajima, N., Tanaka, A., Kotani, H., Nomura, N. and Ohara, O.,
 Prediction of the coding sequences of unidentified human genes.
 XII. The complete sequences of 100 new cDNA clones from brain which
 code for large proteins in vitro
 JOURNAL DNA Res. 5 (6), 355-364 (1998)
 MEDLINE 99156230
 PUBMED 10048495
 REFERENCE 2 (bases 1 to 5414)
 Ohara, O., Suyama, M., Kikuno, R., Nagase, T. and Ishikawa, K.,
 Direct Submission
 TITLE Submitted (02-DEC-1998) Osamu Ohara, Kazusa DNA Research Institute,
 Laboratory of DNA Technology, Yana 1532-3, Kisarazu, Chiba
 292-0812, Japan (E-mail:cdnaiffo@kazusa.or.jp, Tel:+81-438-52-3913,
 Fax:+81-438-52-3914)
 COMMENT On May 17, 2001 this sequence version replaced gi:4240186.
 FEATURES
 Location/Qualifiers
 1..5414
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="fh04363"
 /sex="male"
 /tissue_type="brain"
 /clone_lib="pBluescriptII SK plus"
 /dev_stage="adult"
 /note="vector:pBluescriptII SK plus. This sequence is
 replaced that of hk05904 cDNA as a representative cDNA
 sequence for KIAA0849"
 gene 1..5414
 /gene="KIAA0849"
 CDS <426..3308
 /product="KIAA0849 protein"
 /protein_id="BAA74872.2"
 /db_xref="GI:14133221"

/translation="YFEYNIITMSSGLMSQEKVTSYPWBERIFYLLIQECVSVDKQOTK
 LLKVPKSGIGVIODRSVGHSRIPSAKGKNQIGLILKLEOPHAFVLFVDEKDVVEINEK
 FTELLLATNGCEERPSLPKNNRSLKGLQIDVGCVKVQLSGSEKPGCVWFRGPDLL
 AERTVSGIFFGVELLSEGRGGQFTDGTQGGKQFQCDDEDCGVFVADLKLELLEDDTA
 LESDYAGPDQTMQVELPLEINSRSLKVGTEISGTVIFCDVLPKESLGYFVGVDV
 DNPIGNWDRFDVQLCSFACVESTILLIINDIIPESVTOERRPKLAFMMSGVGDGK
 SSNNKPKATGTSIDPNRNSRSLFTLNGSVSDSQSQSKNTWYIDVADDPKSL
 TEISTDFRSPPPLQPPNVSLTTENRPHSLPFLSKMPTNKGSIHSPLSLSAOSVM
 BELTAPVQESPPPLAMPNGSHGLEVSLAEVKENPPFVGIVIRWICQPPGLNEVLGL
 BLEDCACCTDGTFRGTRIFTALKKALFKLKSPPSRFASLPQVSNQIERCNSLA
 FGGYSEVEVENTPMKEGTELEIMIGKKGIQGHNSCYLSTLFLCFASFSSVLDTV
 LLRPKKNDEYVYSETQLLRTIIVNPLRIYGVCAIKIMKLEKILKLEFAASGFTSE
 EKDPDEFNILLFPHILRVEPLLKIRSAQGVQDQCYFOIFMEKNEKVGVTYQQLLEW
 SPFNSNLKFAEAPSLIIQMPRFKQDFLKKIFPSLEINIITDLEDTPRQCRICGL
 AMYECRECYDDPDISAGIKOPCKTCNTQVHLHPKRLNHNPNVSLPKDLPDWRHGG
 CIPQNMELFACIETSHYVAFVYKDDSAWLFDFSDNADRGQNGFNIFQVITPCP
 EVGEYLMKSLDLHSLDSRRITGGCARRLLCDAYCMYQSPMTSLYK"

ORIGIN

Alignment Scores:
 Pred. No.: 0 Length: 5414
 Score: 4983.00 Matches: 948
 Percent Similarity: 99.48% Conservative: 0
 Best Local Similarity: 99.48% Mismatches: 1
 Query Match: 98.99% Indels: 4
 DB: 9 Gaps: 4
 US-09-671-687A-3 (1-949) x AB020656 (1-5414)
 QY 1 MetSerSerGlyLeuTTPSerGlnGluLysValThrSerProTyrTTPGluGluArgIle 20
 Db 447 ATGAGTTTCAAGCTTATGAGCCCAAGAAAAGTCACTTCACTCCCTACTGGGAAGCGGATT 506
 QY 21 PheTyrLeuLeuGlnGluCysSerValThrAspLysGlnThrGlnLysLeuLeuLys 40
 Db 507 TTTTACTTCTCTTCAAGAATGCAGCGTTACAGACAAACAAACAAACAAAGCTCTCTAAA 566
 QY 41 ValProLysGlySerIleGlyGlnTyrIleGlnAspArgSerValGlyHisSerArgIle 60
 Db 567 GTACCGAAGGGAAGTATAGGACAGTATATTCAGAGATCGTTCTGTGGGGCATTTCAAGATT 626
 QY 61 ProSerAlaLysGlyLysLysAsnGlnIleGlyLeuLysIleLeuGluGlnProHisAla 80
 Db 627 CCTTCTGCAAAAGCGCAAGAAAATCAGATGGATTAAATAATTTCTAGAGCAACCTCATGCA 686
 QY 81 ValLeuPheValAspGlu---AspValValGluIleAsnGluLysPheThrGluLeuLeu 99
 Db 687 GTTCTCTTTGTTGATGAAAAGGATGTTGTAGAGATAAATAAGTTCACAGAGTTACTTT 746
 QY 100 LeuAlaIleThrAsnGlyGluArgPheSerLeuPheLysAsnArgAsnArgLeuSer 119
 Db 747 TTGGCAATTTACCAATTTGTGAGGAGAGGTTTCAGCCTGTTTAAATAACAGAACACTAAGT 806
 QY 120 LysGlyLeuGlnIleAspValGlyCysProValLysValGlnLeuArgSerGlyGluGlu 139
 Db 807 AAAGGCTTCCAATAGACGTGGGCTGTCTCTGTAAGTACAGCTGAGATCTGGGGAGAA 866
 QY 140 LysPheProGlyValValArgPheArgGlyProLeuLeuAlaGluArgThrValSerGly 159
 Db 867 AAATTTCTCTGGAGTTGTACGCTTCAGAGACCCCTGTTAGCAGAGAGGACAGCTCTCCGA 926
 QY 160 IlePhePheGlyValGluLeuLeuGluGlyArgGlyGlnGlyPheThrAspGlyVal 179
 Db 927 ATATTCTTTGGAGTTGAATTTGTGGAAGAAGTCTGTGGTCAAGGTTTCTATGACGGGGTG 986
 QY 180 TyrGlnGlyLysGlnLeuPheGlnCysAspGluAspCysGly---PheValAlaLeuAsp 198
 Db 987 TACCAGGGAACAGCTTTTTCAGTGTGATGAAGATTGTGGCGTGTGTTGTTGATTGGAC 1046
 QY 199 LysLeuGluLeuIleGluAspAspThrAlaLeuGluSerAspTyrAlaGlyProGly 218
 Db 1047 AAGCTAGAACTCATAGAACATGATGACACTGCATCTGGAAAGTATTACGAGGCTCTCGG 1106

QY 219 AspThrMetGlnValGluLeuProProLeuGluIleAsnSerArgValSerLeuLysGly 238
DB 1107 GACAAATGCGAGTCGAACTTCCTCTTGGAAATAAACTCCAGAGTTCTCTTTGAAGGTT 1166
QY 239 GlyGluThrIleGluSerGlyThrValIlePheCysAspValLeuProGlyLysGluSer 258
DB 1167 GGAGAAACAATAGNATCTGGACAGTTATATCTGTGATGTTTTGCCAGGAAAGAAGC 1236
QY 259 LeuGlyTyrPheValGlyValAspMetAspAsnProIleGlyAsnTrpAspGlyArgPhe 278
DB 1227 TTAGGATATTTGTTGGTGGACATGGATAACCTATTGGCAACTGGGATGGAGATT 1286
QY 279 AspGlyVal---LeuCysSerPheAlaCysValGluSerThrIleLeuLeuHisIleAsn 297
DB 1287 GATCGAGTCGACGCTTTGTAGTTTTGCGTGTGTTGAAAGTACAATTTCTATTGCACATCAAT 1346
QY 298 AspileileProGluSerValThrGlnGluArgArgProProLysLeuAlaPheMetSer 317
DB 1347 GATATCATCCAGAGAGTGTGACCGAAGAGGAGGCTCCCAAACCTTGCCCTTTATGTCA 1406
QY 318 ArgGlyValGlyAspLysGlySerSerHisAsnLysProLysAlaThrGlySerThr 337
DB 1407 AGAGGTGTTGGGACAAAGGTTTCATCCAGTCATAATAACCAAGGCTACAGGATCTACC 1466
QY 338 SerAspProGlyAsnArg---ArgSerGluLeuPheTyrThrLeuAsnGlySerSerVal 356
DB 1467 TCAGACCTCGAAATAGAAACAGATCTGAAATATTTTATATACCTTAAATGGGCTCTCTGTT 1526
QY 357 AspSerGlnProGlnSerLysSerLysAsnThrTrpTyrIleAspGluValAlaGluAsp 376
DB 1527 GACTCACAAACCAATCCAAATCAAAATAACATGATGTTGATGAAGTTGCAGAAAGAC 1586
QY 377 ProAlaLysSerLeuThrGluIleSerThrAspPheAspArgSerSerProProLeuGln 396
DB 1587 CCTGCAAAATCTCTACAGAGATATCTACAGACTTTGACCGTCTTCCACCACCACTCCAG 1646
QY 397 ProProProValAsnSerLeuThrThrGluAsnArgPheHisSerLeuProPheSerLeu 416
DB 1647 CCTCCTCCTGTGAACCTCACTGACCACCGAGAACAGATTCCACTCTTTTACCATTCACTC 1706
QY 417 ThrLysMetProAsnThrAsnGlySerIleGlyHisSerProLeuSerLeuSerAlaGln 436
DB 1707 ACCAAGATGCCAAATACCAATGAAAGTATTGGCCACAGTCCACTTCTCTGTGAGCCAG 1766
QY 437 SerValMetGluGluLeuAsnThrAlaProValGlnGluSerProProLeuAlaMetPro 456
DB 1767 TCTGTAATGGAAGAGCTAAACACTGCACCCGTCAGAGAGTCCACCTTGGCCATGCCT 1826
QY 457 ProGlyAsnSerHisGlyLeuGluValGlySerLeuAlaGluValLysGluAsnProPro 476
DB 1827 CCTGGGAACTCATGGTCTAGAAAGTGGGCTCATTTGGCTGAAGTTAAGGAAACCTCCT 1886
QY 477 PheTyrGlyValIleArgTrpIleGlyGlnProProGlyLeuAsnGluValLeuAlaGly 496
DB 1887 TTCTATGGGTAATCCGTTGGATCGGTGAGCCACCAAGAGTGAATGAAGTGCCTCGCTGA 1946
QY 497 LeuGluLeuGluAspGluCysAlaGlyCysThrAspGlyThrPheArgGlyThrArgTyr 516
DB 1947 CTGGAACCTGGAAGATGAGTGTGAGGCTGTACCGATGGAACTTCAGAGGACTCCGGTAT 2006
QY 517 PheThrCysAlaLeuLysLysAlaLeuPheValLysLeuLysSerCysArgProAspSer 536
DB 2007 TTCACCTGTGCTTGAAGAAGGCGCTGTTGTGAACTGAAGAGCTGCAGGCTTGACTCT 2066
QY 537 ArgPheAlaSerLeuGlnProValSerAsnGlnIleGluArgCysAsnSerLeuAlaPhe 556
DB 2067 AGGTTTGATCAATTGACGCGGTTTCCAAATCAGATTGAGCGCTGTAATCTCTTAGCATTT 2126
QY 557 GlyGlyTyrLeuSerGluValValGluGluAsnThrProProLysMetGluLysGluGly 576
DB 2127 GGAGGCTACTTAAGTAGAGTAGAGAAATACTCCACCAAAATGGAAGAAAGGAGGC 2186
QY 577 LeuGluIleMetIleGlyLysLysGlyIleGlnGlyHisTyrAsnSerCysTyrLeu 596

DB 2187 TTGGAGATAATGATGGGAAGAAAGGCATCCAGGGTCATTACAATTTCTTGTACTTTA 2246
QY 597 AspSerThrLeuPheCysLeuPheAlaPheSerSerValLeuAspThrValLeuLeuArg 616
DB 2247 GACTCAACCTTATTTCTGCTTATTTGCTTTTAGTTCTGTCTGGACACTGTGTACTTAGA 2306
QY 617 ProLysGluLysAsnAspValGluTyrTyrSerGluThrGlnGluLeuLeuArgThrGlu 636
DB 2307 CCCAAAGAAAGAACGATGTAGATAATATATAGTGAACCCCAAGAGCTACTGAGGACAGAA 2366
QY 637 IleValAsnProLeuArgIleTyrGlyTyrValCysAlaThrLysIleMetLysLeuArg 656
DB 2367 ATTGTTAATCTCTCGAGAAATATATGATATGTGTGCGCACAAAATATATGAACCTGAGG 2426
QY 657 LysIleLeuGluLysValGluAlaSerGlyPheThrSerGluGluLysAspProGlu 676
DB 2427 AAAATACTTGAAGAGTGGAGGCTGCATCAGATTTCCTCTGAAAGAAAAGATCCTGAG 2486
QY 677 GluPheLeuAsnIleLeuPheHisIleLeuArgValGluProLeuLeuLysIleArg 696
DB 2487 GAATTTCTGAATATTTCTGTTTCATCAATATTTAAGGTTAGAACCTTTGCTAAAAATAGA 2546
QY 697 SerAlaGlyGlnLysValGlnAspCysTyrPheTyrGlnIlePheMetGluLysAsnGlu 716
DB 2547 TCAGCAGGTCAAAAGGTACAAAGATTGTTACTTCTATCAAAATTTTATGCAAAAAAATGAG 2606
QY 717 LysValGlyValProThrIleGlnGlnLeuLeuGluTrpSerPheIleAsnSerAsnLeu 736
DB 2607 AAAGTTGGCGTCCCAAAATTCAGCAGTTGTAGATGGTCTTTTATCAACAGTAGAACCTG 2666
QY 737 LysPheAlaGluAlaProSerCysLeuIleGlnMetProArgPheGlyLysAspPhe 756
DB 2667 AAATTCAGAGGACCATCATGTCGTATATTAGATGCGCTCGATTTGGAAAAGACTTT 2726
QY 757 LysLeuPheLysLysIlePheProSerLeuGluLeuAsnIleThrAspLeuLeuGluAsp 776
DB 2727 AAACATAATTAATAATTTTCTCTCTGGAATTAATAATAACAGATTTACTTGAAGAC 2786
QY 777 ThrProArgGlnCysArgIleCysGlyGlyLeuAlaMetTyrGluCysArgGluCysTyr 796
DB 2787 ACTCCACAGAGTCCCGGATATGTGGAGGGCTTGCAATGTATGAGTGTAGAGAATGCTAC 2846
QY 797 AspAspProAspIleSerAlaGlyLysIleLysGlnPheCysLysThrCysAsnThrGln 816
DB 2847 GACATCCGACATCTCAGCTGGAATAATCAAGCAGTTTTGTAAACCTGCAACACTCAA 2906
QY 817 ValHisLeuHisProLysArgLeuAsnHisLysTyrAsnProValSerLeuProLysAsp 836
DB 2907 GTCCACCTTCATCCGAAGAGCTGAATCATATAAATAATAACCCAGTGTCACTTCCCAAGAC 2966
QY 837 LeuProAspTrpAspTrpArgHisGlyCysIleProCysGlnAsnMetGluLeuPheAla 856
DB 2967 TTACCCGACTGGGACTGGAGACACGGCTGCATCCCTTGCCAGAATATGAGATTTATTGCT 3026
QY 857 ValLeuCysIleGluThrSerHisTyrValAlaPheValLysTyrGlyLysAspSer 876
DB 3027 GTTCTCTGCATAGAAACAACCCACTATGTGCTTTTGTGAAGTATGGGAAGGACGATCT 3086
QY 877 AlaTrpLeuPhePheAspSerMetAlaAspArgAspGlyGlyGlnAsnGlyPheAsnIle 896
DB 3087 GCCTGGCTCTCTTTGACAGCATGGCCGATCGGATGGTGGTGCAGATGGCTTCAACATT 3146
QY 897 ProGlnValThrProCysProGluValGlyGlyLeuLysMetSerLeuGluAspLeu 916
DB 3147 CCTCAAGTCAACCCATGCCCAGAAAGTAGGAGAGTACTTGAAGATGTCTCTGGAAGACCTG 3206
QY 917 HisSerLeuAspSerArgArgIleGlnGlyCysAlaArgArgLeuLeuCysAspAlaTyr 936
DB 3207 CATTCCTTGGACTCCAGGAGATCCAAAGGCTGTGCAGAAAGACTGCTTTGTGATGCATAT 3266
QY 937 MetCysMetTyrGlnSerProThrMetSerLeuTyrLys 949

Db	3267	ATGTGCATGACCAGAGTCCAAACAAATGAGTTTGTACAAA	3305
RESULT 3			
HSA250014			
LOCUS	HSA250014	5371 bp	mRNA
DEFINITION	Homo sapiens mRNA for Familial Cylindromatosis cyld gene.		PRI 16-JUN-2000
ACCESSION	AJ250014		
VERSION	AJ250014.1	GI:8250235	
KEYWORDS	cyld gene; Familial Cylindromatosis.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE			
AUTHORS	Bignell, G.R., Brown, C., Biggs, P.J., Lakhani, S.R., Jones, C., Hansen, J., Blair, B., Hofmann, B., Siebert, R., Turner, G., Evans, D.G., Schander-Stumpel, C., Beemer, F.A., Van Den Ouweland, A., Hallley, D., Delpech, B., Cleveland, M.G., Leigh, I., Leisti, J., Rasmussen, S., Wallace, M.R., Fenske, C., Banerjee, P., Oiso, N., Chaggar, R., Merrett, S., Leonard, N., Huber, M., Hohl, D., Chapman, P., Burn, J., Swift, S., Smith, A., Ashworth, A. and Stratton, M.R.		
TITLE	Identification of the familial cylindromatosis tumour-suppressor gene		
JOURNAL	Nat. Genet. 25 (2), 160-165 (2000)		
MEDLINE	20296617		
PUBMED	10835629		
REFERENCE	2 (bases 1 to 5371)		
AUTHORS	Stratton, M.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-NOV-1999) Stratton M.R., Cancer Genetics, Institute of Cancer Research, 15 Cotswold Rd, Sutton., Surrey. SM2 5NG., UNITED KINGDOM		
FEATURES			
source	Location/Qualifiers		
	1..5371		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/chromosome="16"		
	/map="16q12 - q13"		
gene	1..5371		
	/gene="cyld"		
CDS	392..3262		
	/gene="cyld"		
	/function="tumour suppressor/recessive oncogene"		
	/note="Familial Cylindromatosis Gene"		
	/evidence=experimental		
	/codon_start=1		
	/db_xref="SPTREMBL:Q9NQC7"		
	/translation="MSSGLWSQEKVTSFYWEERIIFYLLEQESVTDKQTKLLKVPKG		
	SIGQVTDRSVGHRIIPSAKGNKQIGLKILPHAVLFVDEKDVVEINEKFTLLILA		
	ITNCERPSLFKNRNLRSKGLQIDVCPVKVQLRSGEKFPGVVFRGRGLAERTVSG		
	IFFGVELLEGSGQFTDVGQKQLQCDCEGCVFVALDKLELIEDDDTALESYAG		
	PGTGMVLEPILNRSVLKVGTEISGTVIFCDVLPGKSLGIFVGMNDNPIGNW		
	DGRFGVQICSFACVESTILLINDIIPALSESVTQERRPPKLFMSRGVGDGSSSH		
	NPKATGTSDFGNRNSLFTLNGSSVDQPSKSNWTYIDEVAEDPAQSLTEIS		
	TFDRSSPPQLPPNRSLLTENRFLSLPFLSLMPTNNGSIHSPLSLSAQSMBELN		
	TAPQESPLAMPNGSHGLVGLAEVKENPPFYGVIRWIGOPPLNEVLAGLEED		
	ECAGCTDGTFRCTKALKKALFVKLSCRPSDFASLPQVSNQIERCNSLFGGY		
	LESVVENTPPMKEGLEIMIGKKGIQGHYNSCYLDTSLFCLFAPSSVLDTVLRP		
	KEKNDEVYSYTOELLRLIVNPLRIYGVYCATKIMKRLKEKVEAASGFTSEBKDP		
	BEFLNLFHILRPLLRKIRAGQVQDQYFIQFMENKXVGVPTIQOLLEWFIN		
	SNLKFAEAPSCILIIOMPRFGKDFLKKIIPSLNLTIDLEDTPRCRCICGLWAME		
	CRECYDDPDI SAGKI KQPKCTNTQVHLHPKLNHNKYNPVSPLKDLPDMDWRHGCTPC		
	ONMELFALCIETSHVAVFVKYKDDSAWLFPDSMDRDGGQNGFNIPQVTPCPVEGE		
	YLKMSLEDDHLSDRRIQGCARRLLCDATMCMIQSPMTSLYK"		
ORIGIN			
Alignment Scores:			
Pred. No.:	0	Length:	5371
Score:	4971.50	Matches:	948

Db 1352 TTTATGTCAAGAGGTGTTGGGACAAAGGTTTCATCCAGTCATATAATAAACCAAGGCTACA 1411
Qy 335 GlySerThrSerAspProGlyAsnArg- --ArgSerGluLeuPheTyrThrLeuAsnGly 353
Db 1412 GGATCTACCTCAGACCCCTGGAATAGAAACAGATCTGAATATATTTATACCTTAAATGGG 1471
Qy 354 SerSerValAspSerGlnProGlnSerLysSerLysAsnThrTrpTyrIleAspGluVal 373
Db 1472 TCTTCTGTTGACTCACAAACACCAATCCAAATCAAAATCAAAATACATGGTATCATTTGAAGTT 1531
Qy 374 AlaGluAspProAlaLysSerLeuThrGluIleSerThrAspPheAspArgSerSerPro 393
Db 1532 GCAGAGACCCCTGCAAAATCTCTTACAGATATCTACAGACTTTGACCGTCTTTCACCA 1591
Qy 394 ProLeuGlnProProValAsnSerLeuThrThrGluAsnArgPheHisSerLeuPro 413
Db 1592 CCACTCCAGCCTCCTCTGTGAACCTCACTGACCACCCAGAACAGATTCCTTTACCA 1651
Qy 414 PheSerLeuThrLysMetProAsnThrAsnGlySerIleGlyHisSerProLeuSerLeu 433
Db 1652 TTCAGTCTCAACCAAGATGCCCAATACCAATGGAAGTATTGGCCACAGTCCACTTCTCTG 1711
Qy 434 SerAlaGluSerValMetGluLeuLeuAsnThrAlaProValGlnGluSerProProLeu 453
Db 1712 TCAGCCAGTCTGTAAATGGAGAGCTAAACACTGCACCCCTGCAAGAGAGTCCACCCCTG 1771
Qy 454 AlaMetProProGlyAsnSerHisGlyLeuGluValGlySerLeuAlaGluValLysGlu 473
Db 1772 GCCATGCTCTGGGAACCTCACATGGTCTAGAGTGGGCTCATTGGCTGAAGTTAAGGAG 1831
Qy 474 AsnProProPheTyrGlyValIleArgTrpIleGlyGlnProProGlyLeuAsnGluVal 493
Db 1832 AACCTCTCTTCTATGGGTAAATCCGTTGGATCGGTGAGTCCAGGACTGAATGAAGTG 1891
Qy 494 LeuAlaGlyLeuGluLeuGluAspGluCysAlaGlyCysThrAspGlyThrPheArgGly 513
Db 1892 CTCGCTGGACTGGAACCTGGAAGATGAGTGTGCAAGGCTGTACGGATGGAACCTTCAGAGGC 1951
Qy 514 ThrArgTyrPheThrCysAlaLeuLysLysAlaLeuPheValLysLeuLysSerCysArg 533
Db 1952 ACTCGGTATTTACCTGTGCTGCTGGAAGGCGCTGTTGTGAACCTGAAGAGCTGCAGG 2011
Qy 534 ProAspSerArgPheAlaSerLeuGlnProValSerAsnGlnIleGluArgCysAsnSer 553
Db 2012 CCTGACTAGGTTTGACTCAITTCAGCCGCTTCCAAATCAGATTCAGCGCTGTAACCTCT 2071
Qy 554 LeuAlaPheGlyGlyTyrLeuSerGluValValGluGluAsnThrProProLysMetGlu 573
Db 2072 TTAGCATTTGGAGGCTACTTAAGTGAAGTAGTAGAAGAAATACTCCACCAAAATGGAA 2131
Qy 574 LysGluGlyLeuGluIleMetIleGlyLysLysGlyIleGlnGlyHisTyrAsnSer 593
Db 2132 AAAGAAGGCTGGAGATAATGATTTGGGAAGAAAGGCAATCCAGGGTCAATTACAATCT 2191
Qy 594 CysTyrLeuAspSerThrLeuPheCysLeuPheAlaPheSerSerValLeuAspThrVal 613
Db 2192 TGTACTTAGACTCAACCTTATTCCTGTTATTTGCTTTAGTTCTGTCTGGACACTGTG 2251
Qy 614 LeuLeuArgProLysGluLysAsnAspValGluTyrTyrSerGluThrGlnGluLeuLeu 633
Db 2252 TTACTTAGACCCAAAGAAAGACGATAGATAATATTAGTGAACCCCAAGAGCTACTG 2311
Qy 634 ArgThrGluIleValAsnProLeuArgIleTyrGlyTyrValCysAlaThrLysIleMet 653
Db 2312 AGGACAGAAATTTTAATCCTCTGAGAATATATGATATGTGTGTGTCACAAAAATATG 2371
Qy 654 LysLeuArgLysIleLeuGluLysValGluAlaSerGlyPheThrSerGluGluLys 673
Db 2372 AAACCTGAGGAAATATCTTGAAGAGGTGGAGCTGCATCAGGATTTACCTCTGAAGAAAA 2431
Qy 674 AspProGluGluPheLeuAsnIleLeuPheHisIleLeuArgValGluProLeuLeu 693
Db 2432 GATCCTGAGGAATCTTGAATATCTTGTTCATCATATTTTAAGGGTAGAACCTTTGCTA 2491

Qy 694 LysIleArgSerAlaGlyGlnLysValGlnAspCysTyrPheTyrGlnIlePheMetGlu 713
Db 2492 AAAATAAGATCAGCAGGTCAAAAGGTCAAGATTGTTACTTCTATCAAAATTTTATGGAA 2551
Qy 714 LysAsnGluLysValGlyValProThrIleGlnGlnLeuLeuGluTrpSerPheIleAsn 733
Db 2552 AAAAATGAGAAAGTTGGCGTTCCCACAATTCAGCAGTTGTTAGAAATGCTTTTATCAAC 2611
Qy 734 SerAsnLeuLysPheAlaGluAlaProSerCysLeuIleIleGlnMetProArgPheGly 753
Db 2612 AGTAACCTGAATTTGAGAGGACCAATCATGTCGTGATTATTCAGATGCCCTCGATTGGA 2671
Qy 754 LysAspPheLysLeuPheLysLysIlePheProSerLeuGluLeuAsnIleThrAspLeu 773
Db 2672 AAAGACTTTAAACTATTATAAAAAATTTTCTCTCTGGAATTTAAATATAACAGATTTA 2731
Qy 774 LeuGluAspThrProArgGlnCysArgIleCysGlyGlyLeuAlaMetTyrGluCysArg 793
Db 2732 CTTGAAGACACTCCACAGACAGTCCGATATGTGGAGGCTTGCAATGTATGAGTGTAGA 2791
Qy 794 GluCysTyrAspAspProAspIleSerAlaGlyLysIleLysGlnPheCysLysThrCys 813
Db 2792 GAATGCTACGACGATCCGACATCTCAGCTGGAAAAATCAAGCAGTTTTGTAAAAACCTGC 2851
Qy 814 AsnThrGlnValHisLeuHisProLysArgLeuAsnHisLysTyrAsnProValSerLeu 833
Db 2852 AACACTCAAGTCCACCTTCATCCGAGAGGCTGAATCATATAATATAACCCAGTGTCACTT 2911
Qy 834 ProLysAspLeuProAspTrpAspTrpArgHisGlyCysIleProCysGlnAsnMetGlu 853
Db 2912 CCCAAAGACTTACCGGACTGGGACTGGAGACACGGCTGCATCCCTTGGCCAGAAATATGGAG 2971
Qy 854 LeuPheAlaValLeuCysIleGluThrSerHisTyrValAlaPheValLysTyrGlyLys 873
Db 2972 TTATTTGCTGTTCTCTGCATAGAAACAAAGCCACTATGTTGCTTTGTGAAGTATGGGAAG 3031
Qy 874 AspAspSerAlaTrpLeuPhePheAspSerMetAlaAspArgAspGlyGlyGlnAsnGly 893
Db 3032 GACGATTCGCTCGGCTCTCTTTTGACGAGCATGCCGATCGGATGGTGTGTCAGAAATGCG 3091
Qy 894 PheAsnIleProGlnValThrProCysProGluValGlyGluTyrLeuLysMetSerLeu 913
Db 3092 TTCAACATTCCTCAAGTCAACCCATGCCAGAGTAGGAGAGTACTTTGAAGATGTCTCTG 3151
Qy 914 GluAspLeuHisSerLeuAspSerArgArgIleGlnGlyCysAlaArgArgLeuLeuCys 933
Db 3152 GAAGACCTGCTATCTCTGGACTCCAGGAGAAATCCAGGCTGTGCACGAGACTGCTTTGT 3211
Qy 934 AspAlaTyrMetCysMetTyrGlnSerProThrMetSerLeuTyrLys 949
Db 3212 GATGCATATATGTGCATGTACCAGAGTCCAACAATGATTTGTACAAA 3259
RESULT 4
BD231207
LOCUS 4527 bp DNA linear PAT 17-JUL-2003
DEFINITION Human cytoskeleton associated proteins.
ACCESSION BD231207
VERSION BD231207.1 GI:33040977
KEYWORDS JP 2002526076-A/9.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 4527)
AUTHORS Lal, P., Tang, T. Y., Yue, H., Hillman, J. L., Bandman, O., Corley, N. C.,
Guegler, K. J., Patterson, C., Azimzai, Y. and Baughn, M. R.
TITLE Human cytoskeleton associated proteins
JOURNAL Patent: JP 2002526076-A 9 20-AUG-2002;
INCYTE PHARMACEUTICALS INC
COMMENT OS Homo sapiens (human)
PN JP 2002526076-A/9
PD 20-AUG-2002

PF 17-SEP-1999 JP 2000574254
 PR 18-SEP-1998 US 60/172226,27-APR-1999 US 60/131321 PI
 PREETI LAL, TOM Y TANG, HENRY YUE, JENNIFER L HILLMAN, OLGA PI
 BANDMAN,

PI NEIL C CORLEY, KARL J GUEGLER, CHANDRA PATTERSON, YALDA AZIMZAI,
 PI MARIAH R BAUGHN
 PC C12N15/09, A61K38/00, A61K45/00, A61P1/16, A61P3/00, A61P3/06 PC
 , A61P3/10, A61P5/00,

PC A61P5/38, A61P7/00, A61P7/06, A61P9/00, A61P9/10, A61P11/06 PC
 , A61P13/00, A61P13/12,
 PC A61P17/06, A61P19/02, A61P19/10, A61P21/04, A61P25/00, A61P25/00,
 A61P25/04,

PC A61P25/14, A61P25/16, A61P25/28, A61P29/00, A61P31/04, A61P31/12,
 PC A61P31/18,
 PC A61P33/00, A61P35/00, A61P35/02, A61P37/00, A61P43/00, C07K14/47,
 PC C07K16/18,

PC C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12P21/02, C12P21/08 PC
 , C12Q1/68, G01N33/15,
 PC G01N33/50, G01N33/53, G01N33/53, G01N33/53, G01N33/566, C12N15/00, C12N5/00,
 PC A61K37/02

CC Incyte ID No: 2363327

FH Key Location/Qualifiers

FT source 1..4527 /organisms='Homo sapiens (human)'.
 FT Location/Qualifiers

1..4527 /organisms='Homo sapiens (human)'.
 /organism='Homo sapiens'
 /mol_type='genomic DNA'
 /db_xref='taxon:9606'

FEATURES

source

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 4527
 Score: 4942.00 Matches: 944
 Percent Similarity: 99.37% Conservative: 0
 Best Local Similarity: 99.37% Mismatches: 2
 Query Match: 98.17% Indels: 5
 DB: 6 Gaps: 4

US-09-671-687A-3 (1-949) x BD231207 (1-4527)

QY 4 GlyLeuTTPSerGlnGluLysValThrSerProTyrTTPGluGluArgIlePheTyrLeu 23

DB 270 GGCATTATGAGCCAGAAAAGTCACTTCACCTACTGGGAAGCGGATTTTACTTTC 329

QY 24 LeuLeuGlnGluCysSerValThrAspLysGlnThrLysLeuLeuLysValProLys 43

DB 330 CTTCTTCAAGATGCAGCGTTACAGACAAACAAACAAAGCTCCTTAAAGTACCGAA- 388

QY 44 GlySerIleGlnTyrIleGlnAspArgSerValGlyHisSerArgIleProSerAla 63

DB 389 GGAAGTATAGGACAGTATATCAAGATCGTTCTGTGGGCAATTCAGGATTCCTCTGCA 448

QY 64 LysGlyLysLysAsnGlnIleGlyLeuLysIleLeuGluGlnProHisAlaValLeuPhe 83

DB 449 AAAGGCAAGAAAATCAGATTGATTAATAATTTCTAGAGCAACCTCATCGAGTTCCTTT 508

QY 84 ValAspGlu---AspValValGluIleAsnGluLysPheThrGluLeuLeuAlaIle 102

DB 509 GTTGATGAAAGGATGTTGTAGAGATAAATGAAAGTTCCACAGATTACTTTTGCAATT 568

QY 103 ThrAsnCysGluGluArgPheSerLeuPheLysAsnArgAsnArgLeuSerLysGlyLeu 122

DB 569 ACCAATTGTGAGGAGAGGTTACGCTGTTTAAAAACAGAAACAGACATAAGTAAAGCCTC 628

QY 123 GlnIleAspValGlyCysProValLysValGlnLeuArgSerGlyGluLysPhePro 142

DB 629 CAAATAGACGTGGGCTGCTCTGTGAAAGTACAGCTCAGATCTGGGGAAGAAAATTCCT 688

QY 143 GlyValValArgPheArgGlyProLeuLeuAlaGluArgThrValSerGlyIlePhePhe 162

DB 689 GGAGTTGTACGCTTCAGAGACCCCTGTTAGCAGAGAGGACAGCTCTCCGGATATTCCTT 748

QY 163 GlyValGluLeuLeuGluGluGlyArgGlyGlnGlyPheThrAspGlyValTyrGlnGly 182

DB 749 GGAGTTGAATTCGTGGAAGAGTCTGTGTCAAGGTTTTCACGTACGGGGGTGTACCAAGG 808

QY 183 LysGlnLeuPheGlnCysAspGluAspCysGly---PheValAlaLeuAspLysLeuGlu 201

DB 809 AAACAGCTTTTTCAGTGTGATGACAGATTGTGCGGTGTTGTTGTCATGTCACCAAGCTAGAA 868

QY 202 LeuIleGluAspAspThrAlaLeuGluSerAspTyrAlaGlyProGlyAspThrMet 221

DB 869 CTCATAGAGATGATGACACTGTCATTGGAAGTGTATAGCGAGGTCCTGGGACACAAATG 928

QY 222 GlnValGluLeuProLeuGluIleAsnSerArgValSerLeuLysGlyGlyGluThr 241

DB 929 CAGGTCGAACCTCTCTTTGGAAATAAATCCAGAGTTCTTTGNAAGTTGGAGAAACA 988

QY 242 IleGluSerGlyThrValIlePheCysAspValLeuProGlyLysGluSerLeuGlyTyr 261

DB 989 ATAGAATCTGGAACAGTTATATCTGTGATGTTTGGCCAGGAAAAGAAAGCTTAGGATAT 1048

QY 262 PheValGlyValAspMetAspAsnProIleGlyAsnTyrAspGlyArgPheAspGlyVal 281

DB 1049 TTTGTTGTTGTGGACATGATAACCTTATTTGCCAATCGGATGGAGATTTGATGGAGTG 1108

QY 282 ---LeuCysSerPheAlaCysValGluSerThrIleLeuLeuHisIleAsnAspIleIle 300

DB 1109 CAGCTTTGTAGTTTTCGTTGTGTAAGTACAATCTTATTGACATCATGATGATCATC 1168

QY 301 ProGluSerValThrGlnGluArgProProLysLeuAlaPheMetSerArgGlyVal 320

DB 1169 CCAGAGAGTGTGACGACGAGAAAGAGGCGCTCCCAAACTTGCTTTATGTCAAGAGGTGTT 1228

QY 321 GlyAspLysGlySerSerHisLysProLysAlaThrGlySerThrSerAspPro 340

DB 1229 GGGGCAAAAGGTTTCATCCAGTCATAATAACAAAGGCTACAGGATCTACCTCAGACCT 1288

QY 341 GlyAsnArg---ArgSerGluLeuPheTyrThrLeuAsnGlySerSerValAspSerGln 359

DB 1289 GGAATATAGAAACAGATCTGAATATTTTATACCTTAAATGGGTCTTCTGTGACTCACAA 1348

QY 360 ProGlnSerLysSerLysAsnThrTyrTyrIleAspGluValAlaGluAspProAlaLys 379

DB 1349 CCACAATCCAAATCAAAAATAACATGGTACATTTGATGAAGTTGCAGAGACCCCTGCAAAA 1408

QY 380 SerLeuThrGluIleSerThrAspPheAspArgSerSerProLeuGlnProProPro 399

DB 1409 TCTCTTACAGAGATATCTACAGACTTTGACCGTTCCTTCCACCACCATCCAGCCCTCCTCT 1468

QY 400 ValAsnSerLeuThrThrGluAsnArgPheHisSerLeuProPheSerLeuThrLysMet 419

DB 1469 GTGAACCTCACTGACCACCGAGAACAGATTCCACTCTTTTACCATTAGTCTCACCAAGATG 1528

QY 420 ProAsnThrAsnGlySerIleGlyHisSerProLeuSerLeuSerAlaGlnSerValMet 439

DB 1529 CCCAATACCAATGGAAGTATTGGCCACAGTCACATTTCTCTGTGACCCAGCTGTAAATG 1588

QY 440 GluGluLeuAsnThrAlaProValGlnGluSerProProLeuAlaMetProProGlyAsn 459

DB 1589 GAAGAGCTAAACACTGACCCGTCCAAGAGAGTCCACCCCTTGGCCATGCTCCTCGGGAAC 1648

QY 460 SerHisGlyLeuGluValGlySerLeuAlaGluValLysGluAsnProPheTyrGly 479

DB 1649 TCACATGGTCTAGAAAGTGGGCTCATTTGGCTGAAGTTAAGAGAACCTCTCTTCTATGGG 1708

QY 480 ValIleArgTyrIleGlyGlnProProGlyLeuAsnGluValLeuAlaGlyLeuGluLeu 499

DB 1709 GTAAATCCGTTGGATCGGTGACCCACAGGACTGAATGAAGTGTCTCGCTGGACTGGAATG 1768

QY 500 GluAspGluCysAlaGlyCysThrAspGlyThrPheArgGlyThrArgTyrPheThrCys 519

DB 1769 GAAGATGAGTGTGAGGCTGTACGGATGGAACCTTCAGAGGCACTCGGTATTTTACCTGT 1828

QY 520 AlaLeuLysLysAlaLeuPheValLysLeuLysSerCysArgProAspSerArgPheAla 539

Db 1829 GCGCTGAAGAGGCGCTGTTGTGTAAGAGCTGCGAGCGCTGACTCTAGGTTTGCA 1888
Qy 540 SerLeuGlnProValSerAsnGlnIleGluArgCysAsnSerLeuAlaPheGlyGlyTyr 559
Db 1889 TCATTGACGCGGTTTCCATCAGATTGAGCGCTGTAACTCTTTAGCATTTGGAGGCTAC 1948
Qy 560 LeuSerGluValValGluGluAsnThrProProLysMetGluLysGluGlyLeuGluIle 579
Db 1949 TTAAGTGAAGTAGTAGAAGAAATACCTCCACAAAATGGAAAAAGAGGCTTGGAGATA 2008
Qy 580 MetIleGlyLysLysGlyIleGlnGlyHisTyrAsnSerCysTyrLeuAspSerThr 599
Db 2009 ATGATTGGGAAGAAAGAGCATCCAGGTCATTACAAATCTTGTGTATTAGACTCAACC 2068
Qy 600 LeuPheCysLeuPheAlaPheSerSerValLeuAspThrValLeuLeuLeuArgProLysGlu 619
Db 2069 TTATTCTGCTATTGCTTTTGTCTGTTCTGACACTGTGTATTCTAGACCCAAAGAA 2128
Qy 620 LysAsnAspValGluTyrTyrSerGluThrGlnGluLeuLeuArgThrGluIleValAsn 639
Db 2129 AAGAACGATCTAGAAATATTATAGTGAACCCAGAGCTACTGAGACAGAAATTTGTAAT 2188
Qy 640 ProLeuArgIleTyrGlyTyrValCysAlaThrLysIleMetLysLeuArgLysIleLeu 659
Db 2189 CCTCTGAGAAATATATGATGTGTGCCACAAAATTTATGAACTGAGGAAATACTT 2248
Qy 660 GluLysValGluAlaSerGlyPheThrSerGluGluLysAspProGluGluPheLeu 679
Db 2249 GAAAAGTGGAGGCTGCATCAGGATTTACCTCTGAAGAAAAGATCTCTGAGGAATCTTG 2308
Qy 680 AsnIleLeuPheHisIleLeuArgValGluProLeuLeuLysIleArgSerAlaGly 699
Db 2309 AATATTCTGTTTCATCATATTTTAAAGGTAGAACCTTTGCTAAATAATGATCAGCAGGT 2368
Qy 700 GlnLysValGlnAspCysTyrPheTyrGlnIlePheMetGluLysAsnGluLysValGly 719
Db 2369 CAAAAGTACAAGATTGTTACTTCTATCAAAATTTTATGAAAAAATGAGAAAGTTGGC 2428
Qy 720 ValProThrIleGlnLeuLeuGluTyrSerPheIleAsnSerAsnLeuLysPheAla 739
Db 2429 GTTCCACAAATTCAGCAGTTGTTAGAAATGCTCTTTATCAACAGTAACCTGAAATTTGCA 2488
Qy 740 GluAlaProSerCysLeuIleIleGlnMetProArgPheGlyLysAspPheLysLeuPhe 759
Db 2489 GAGCACCATCATGCTCGATTATTCAGATCCCTGATTTGGAAGAAAGCTTTAAACTATT 2548
Qy 760 LysLysIlePheProSerLeuGluLeuAsnIleThrAspLeuLeuGluAspThrProArg 779
Db 2549 AAAAAAATTTTCTCTCTCGGAATTAATATAACAGATTACTTGAAGACACTCCAGA 2608
Qy 780 GlnCysArgIleCysGlyGlyLeuAlaMetTyrGluCysArgGluCysTyrAspAspPro 799
Db 2609 CAGTGGCGGATATGTGGAGGCTTGCAATGTATGAGTGTAGAAATGCTACGACGATCCG 2668
Qy 800 AspIleSerAlaGlyLysIleLysGlnPheCysLysThrCysAsnThrGlnValHisLeu 819
Db 2669 GACATCTCAGCTGGAAAAATCAAGCAGTTTGTGTAACACCTGCAACACTCAAGTCCACCT 2728
Qy 820 HisProLysArgLeuAsnHisLysTyrAsnProValSerLeuProLysAspLeuProAsp 839
Db 2729 CATCCGAAGAGGCTGAATCAATAATAATAACCCAGTGTCACTTCCCAAGACTTACCCGAC 2788
Qy 840 TrpAspTrpArgHisGlyCysIleProCysGlnAsnMetGluLeuPheAlaValLeuCys 859
Db 2789 TGGGACTGGAGACACGCGCTGCATCCCTTGCAGAAATATGAGTATTTGCTGTCTCTGC 2848
Qy 860 IleGluThrSerHisTyrValAlaPheValLysTyrGlyLysAspSerAlaTrpLeu 879
Db 2849 ATAGAAACAAGCCACTATGTGCTTTGTGAAGTATGGGAAGACGAGTCTGCTGCTC 2908
Qy 880 PhePheAspSerMetAlaAspArgAspGlyGlnAsnGlyPheAsnIleProGlnVal 899

Db 2909 TTCTTTGACAGCATGGCCGATCGGATGGTGCAGATGGCTCAACATTCCTCAAGTC 2968
Qy 900 ThrProCysProGluValGlyGluTyrLeuLysMetSerLeuGluAspLeuHisSerLeu 919
Db 2969 ACCCATGCCCAGAAAGTAGGAGAGTACTTGAAGATGTCTCTGGAAGACCTGCATTCCTG 3028
Qy 920 AspSerArgArgIleGlnGlyCysAlaArgArgLeuLeuCysAspAlaTyrMetCysMet 939
Db 3029 GACTCCAGGAGAAATCCAAAGGCTGTGCACGAAGACTGCTTTGTGATGCATATATGTGCATG 3088
Qy 940 TyrGlnSerProThrMetSerLeuTyrLys 949
Db 3089 TACCAGAGTCCAAATGAGTTTGTACAAA 3118
RESULT 5
AK122389 4314 bp mRNA linear ROD 15-MAR-2003
LOCUS Mus musculus mRNA for mKIAA0849 protein.
DEFINITION AK122389
ACCESSION AK122389
VERSION 1 GI:28972434
KEYWORDS FLI CDNA.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Okazaki,N., Kikuno,R., Ohara,R., Inamoto,S., Aizawa,H., Yuasa,S., Nakajima,D., Nagase,T., Ohara,O. and Koga,H.
TITLE Prediction of the coding sequences of mouse homologues of KIAA gene: II. The complete nucleotide sequences of 400 mouse KIAA-homologous cDNAs identified by screening of terminal sequences of cDNA clones randomly sampled from size-fractionated libraries
JOURNAL DNA Res. 10, 35-48 (2003)
REFERENCE 2 (bases 1 to 4314)
AUTHORS Okazaki,N., Kikuno,R., Nagase,T., Ohara,O. and Koga,H.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-2003) Hisashi Koga, Kazusa DNA Research Institute, Laboratory for Genome Informatics; 2-6-7 Kazusa-kamatori, Kisarazu, Chiba 292-0818, Japan (E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
COMMENT The CREATE program supported by Japan science and technology corporation; cDNA full insert sequencing: Kazusa DNA Research Institute; cDNA library construction, clone selection and 5'- & 3'-end one pass sequencing.
FEATURES
Location/Qualifiers
1. .4314
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="mbh04492"
/tissue_type="brain"
/dev_stage="adult"
/note="vector:modified pBC SK+"
gene 1. .4314
/gene="mKIAA0849"
CDS <197..3196
/gene="mKIAA0849"
/note="CDS is predicted by in silico analysis. Start codon is not identified."
/codon_start=1
/evidence=not_experimental
/product="mKIAA0849 protein"
/protein_id="BAC65671.1"
/db_xref="GI:28972435"

Db	2195	CCCAAAGAGAAGTAATGATATAGAGTATTACAGTGAGACTCAGGAGCTACTCAGGACACAG	2254
QY	637	IleValAsnProLeuArgIleTyrGlyTyrValCysAlaThrLysIleMetLysLeuArg	656
Db	2255	ATAGTCAATCCTCTGAGAATATATGATATGTGTGTGCCAAGAATATTATGAACCTGAGG	2314
QY	657	LysIleLeuGluLysValGluAlaAaSerGlyPheThrSerGluGluLysAspProGlu	676
Db	2315	AAAATACTTGAAAAGTTGAGGCTGCATCAGGTTTACCTCTCGAAGAAAAGATCTCTGAA	2374
QY	677	GluPheLeuAsnIleLeuPheHisIleLeuArgValGluProLeuLeuLysIleArg	696
Db	2375	GAATTCCTTAATATCCTGTTTCATGATATTTTAAGGCTTGAACCATTTGTTAAAAATAAGA	2434
QY	697	SerAlaGlyGlnLysValGlnAspCysTyrPheTyrGlnIlePheMetGluLysAsnGlu	716
Db	2435	TCAGCAGGTCAAAAAGTTCAAGACTGTAACTTCTATCAAAATTTTATGGAAAAAATGAG	2494
QY	717	LysValGlyValProThrIleGlnGlnLeuLeuGluTrpSerPheIleAsnSerAsnLeu	736
Db	2495	AAAGTTGGAGTACCCACAATTCAGCAGTTATTAGAATGGTCTTTTATCAACAGCAACCTG	2554
QY	737	LysPheAlaGluAlaProSerCysLeuIleIleGlnMetProArgPheGlyLysAspPhe	756
Db	2555	AAATTTGCAGAGCCACCATCTGCTTGAATTATCCAGATGCTCGGTTTGGAAAAAGACTTT	2614
QY	757	LysLeuPheLysIlePheProSerLeuGluLeuAsnIleThrAspLeuLeuGluAsp	776
Db	2615	AAACTATTTTAAAAAAATTTTCTCCCTCGGAATTAATAATACAGATTTACTTGAAGAC	2674
QY	777	ThrProArgGlnCysArgIleCysGlyGlyLeuAlaMetTyrGluCysArgGluCysTyr	796
Db	2675	ACTCCAGCAGTCGCCCATCTGTGGAGACTTCGCGATGACGAGTGCAGGAGTGCTAT	2734
QY	797	AspAspProAspIleSerAlaGlyLysIleLysGlnPheCysLysThrCysAsnThrGln	816
Db	2735	GACGATCCGACATCTCAGCTGGAGAGATCAACAGCTTCTGTAAAGACCTGCAGCACTCAG	2794
QY	817	ValHisLeuHisProLysArgLeuAsnHisLysTyrAsnProValSerLeuProLysAsp	836
Db	2795	GTTCACTTCATCCCAAGAGGTTGAATCATCTTATCATCATCCAGTATCACTTCCCAAGAC	2854
QY	837	LeuProAspTrpAspTrpArgHisGlyCysIleProCysGlnAsnMetGluLeuPheAla	856
Db	2855	TTGCCTGACTGGGACTGGAGACACGGCTGCATCCCTGTCAGAAGATGGAGTTATTGCT	2914
QY	857	ValLeuCysIleGluThrSerHisTyrValAlaPheValLysTyrGlyLysAspAspSer	876
Db	2915	GTTCTCTGATAGAAACGAGCCACTATGTGCTTTTGTGAAGTATGGGAAGATGACTCT	2974
QY	877	AlaTrpLeuPhePheAspSerMetAlaAspArgAspGlyGlnAsnGlyPheAsnIle	896
Db	2975	GCCTGCTTTCTTTTACAGCATGGCGGATCGAGATGGTGGTTCAGAATGGCTTCAACATT	3034
QY	897	ProGlnValThrProCysProGluValGlyGluTyrIleuLysMetSerLeuGluAspLeu	916
Db	3035	CCACAAGTCACGCCCTGCCAGAAAGTGGGAGAGTACTTGAAGATGTCTCTGGAGGACCTG	3094
QY	917	HisSerLeuAspSerArgGlnGlnGlyCysAlaArgArgLeuLeuCysAspAlaTyr	936
Db	3095	CACCTCTTGGACTCCAGAAGATTCAAGGCTGTGCCGCGAGACTCTTTTCCGATGCATAC	3154
QY	937	MetCysMetTyrGlnSerProThrMetSerLeuTyrLys	949
Db	3155	ATGTGATGTACACAGATCCAAACCATGAGCCTGTACAAA	3193
RESULT	6		
BC042438			
LOCUS			
DEFINITION	BC042438	4501 bp mRNA linear ROD 21-OCT-2003	
ACCESSION	BC042438	Mus musculus cylindromatosis (turban tumor syndrome), mRNA (cDNA clone MGC:25429 IMAGE:398377), complete cds.	

VERSION
KEYWORDS
SOURCE

ORGANISM
Mus musculus (house mouse)

Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 4501)

REFERENCE
AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullen, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.C., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, N.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A., Young, A.C., Shevchenko, Y.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Green, E.D.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

TITLE

human and mouse cDNA sequences

JOURNAL
MEDLINE
PUBMED

22388257
12477932

REFERENCE
AUTHORS

2 (bases 1 to 4501)
Direct Submission
Submitted (02-JAN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapsb-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 30 Row: g Column: 3.

FEATURES
source

1. .4501
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:25429 IMAGE:3983771"
/tissue_type="Mammary tumor. Metallothionien-TGF alpha
model. 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCI CGAP_Mam1"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
/genes="Cylid"
1. .4501
/notes="synonyms: C130039D01Rik, 2900009M21Rik, EAC, CDWT,
CYLD1"
/db_xref="LocusID:74256"
/db_xref="MGI:1921506"

gene

CDS	176..3034 /codon_start=1 /product="Riken cDNA 2010013M14" /protein_id="AAH42438.1" /db_xref="GI:27503671" /db_xref="J:locusID:74256" /translation="MSSGLWSQEKVTSPIYWEERIFVYLLQBCSVTDKQTKLLKVPKG SIGVYIDRSVGHSRVPTKGNQIGLKEIQHAFVLFVDEKDVVEINKEFTELLIA ITNCEERLSLPRNRLSKGLQVDGSPVKQLRSGBEKFFPCVFRFRGLPELTVSG IFFGVLEEGRGOGFTDGVYGGKQLFCODEDCGVFVALDKLELIEDDDNGLESDPAG PGDTQVPPPLEINSRVLSLVKRGSTESGTVIFCDVLPKESLGFVFGVDMDNPIGNW DGRFDGQCSFASVESTILLIINDIIPDSVTOERRPPKLAFMGRGVGDGSSSHNKP KVTGSDPGSRNRSSELYTLNGSSVDQSQSKNPWYIDEVAEDPAKSLTEMSSDFG HSSPPPOPSMNSLSRPNFSLTKMPNTNGSMAHSPLSLSVQVMGELNSTPV QESPLPLTSSGNHGLEVLAENKPNPFYVIRWIGQPPGLSDVLAGLEDEECAG CTDGTFRGTRYFTCAKALKALFVKLSKCRPSRPFASLPVSNQIERCNSLAFGGYLSEV VEENTPQWKEGLEIMIGKKKGIOGHYNSCYLSTLFCLEFASALDVTILLRPKERN DIETYSFEQLREIYNPLRIYGVYCATIKMKILKELEVAASGFTSEEDKPEEFL NILEFDILRVPLLRKIRSGAKQVDCNFYQIFMEKNKGVPTTQQLLEWFSFINSLK FAEAPSCILIOMPFRGDKFLPKIFPSLELNIITDLEDTPQCRICGLLAMEYECREB YDDPDISAGTKKQCKTCTOVHLHPRLNHSYHPVSLPKDLPDMDWPHGCI PCQKME LPAVLCIETSHYFAVKYKDDSAWLFPDSDMADRGGQGNIPQVTPCPPEVGEYLKM SLEDLSLDSRRIQGCARLLCDAVMCYQSPMTSLYK" 644..766 /note="CAP GLY; Region: CAP-Gly domain. CAP stands for cytokeleton-associated proteins" /db_xref="CDD:pfam01302"	Db	596	AAATTTCCAGGAGTGTACGCTTCAGAGACCTTTATTAGCGGAGAGACGGTGTCCGGG	655
		Qy	160	IlePhePheGlyValGluLeuLeuGluGluGlyArgGlyGlnGlyPheThrPheGlyVal	179
		Db	656	ATTTCTTTGGAGTAGAATATTGGAAGAAGGTCGTGGTCAAGGTTTACCGATGGGTA	715
		Qy	180	TyrGlnGlyLysGlnLeuPheGlnCysAspGluAspCysGly---PheValAlaLeuAsp	198
		Db	716	TATCAAGGAAGCAGCTTTTCCAGTGTGATGAGGACTGTGGCGTGTTCGTTCGATTGGAC	775
		Qy	199	LysLeuGluLeuIleGluAspAspThrAlaLeuGluSerAspTyrAlaGlyProGly	218
		Db	776	AAACTAGAACTATTAGAAGATGATGACAAATGATGGAAGTATTTTCAGGCCCCAGGA	835
		Qy	219	AspThrMetGlnValGluLeuProProLeuGluIleAsnSerArgValSerLeuLysGly	238
		Db	836	GATACAATGCAAGTTGAACCTCCCTCTTTGGAAATAAATCCAGAGATTTCTTTGAAGGTT	895
		Qy	239	GlyGluThrIleGluSerGlyThrValIlePheCysAspValLeuProGlyLysGluSer	258
		Db	896	CGAAGAAAGTACAGAATCTGGAACAGTATATTCTGTGATGTTTTTACCAGAAAAGAGT	955
		Qy	259	LeuGlyTyrPheValGlyValAspMetAspAsnProIleGlyAsnTrpAspGlyArgPhe	278
		Db	956	CTAGGATATTTTGTGGTGTGGACATGATACCCCTATTGGCAACTGGGATGGAAGGTTT	1015
		Qy	279	AspGlyVal---LeuCysSerPheAlaCysValGluSerThrIleLeuLeuHisIleAsn	297
		Db	1016	GATGGAGTACAGCTCTGTAGTTTTTGCAAAGTGTGTAAGATGACAAATTCCTCTGCACATCAAT	1075
		Qy	298	AspIleIleProGluSerValThrGlnGluArgArgProProLysLeuAlaPheMetSer	317
		Db	1076	GACATCATCCAGATAGCTGACAGAAAGAGGCGCTCCCAAACTTCCTTTAATGTCA	1135
		Qy	318	ArgGlyValGlyAspLysGlySerSerHisAsnLysProLysAlaThrGlySerThr	337
		Db	1136	AGAGGTGTAGGTGACAAAGTTTCATCTAGTCATAATAACCAAGGTTACAGGATCTACC	1195
		Qy	338	SerAspProGlyAsnArg---ArgSerGluLeuPheTyrThrLeuAsnGlySerSerVal	356
		Db	1196	TCAGACCCCTGGAAAGTAGAAACAGATCTGAATTAATTTATACCTTAAATGGGTCATCTGT	1255
		Qy	357	AspSerGlnProGlnSerLysSerLysAsnThrTrpTyrIleAspGluValAlaGluAsp	376
		Db	1256	GACTCACAA---CAATCCAAAGTCCAAATAATCCATGATGATGAAAGTTCAGAGAC	1312
		Qy	377	ProAlaLysSerLeuThrGluIleSerThrAspPheAspArgSerSerProProLeuGln	396
		Db	1313	CCTGCAAAAGTCACTTACAGAGATGCTTCGGACTTCGGACATTCATCTCTCCACCGCAG	1372
		Qy	397	ProProProValAsnSerLeuThrThrGluAsnArgPheHisSerLeuProPheSerLeu	416
		Db	1373	CCTCCTTCCATGAACCTCCTTGTCTAGCAGAGAACAGATTCATCTCTTACCTTCAGCCTG	1432
		Qy	417	ThrLysMetProAsnThrAsnGlySerIleGlyHisSerProLeuSerLeuSerAlaGln	436
		Db	1433	ACAAAGATGCCAAATACATAATGCAGCATGGCTCATAGTCCACTCTCTCTGTGAGTCAG	1492
		Qy	437	SerValMetGluGluLeuAsnThrAlaProValGlnGluSerProProLeuAlaMetPro	456
		Db	1493	TCTGTGATGGGGAGCTGNACAGACACCTGTCAGAGAGTCCACCTTGCCATCTCT	1552
		Qy	457	ProGlyAsnSerHisGlyLeuGluValGlySerLeuAlaGluValLysGluAsnProPro	476
		Db	1553	TCTGGGAATGCACACGGGCTAGAGGTGGGCTCACTGGCTGAAGTAAAGAAACCCCGC	1612
		Qy	477	PheTyrGlyValIleArgTrpIleGlyGlnProProGlyLeuAsnGluValLeuAlaGly	496
		Db	1613	TTCTATGGGGTTATCCGTTGGATTGGCCAGCCAGGCGCTCAGTACGCTGTAGTGGGA	1672
		Qy	497	LeuGluLeuGluAspGluCysAlaGlyCysThrAspGlyThrPheArgGlyThrArgTyr	516

ORIGIN

Alignment Scores:	
Pred. No.:	0
Score:	4718.50
Percent Similarity:	96.54%
Best Local Similarity:	94.23%
Query Match:	10
DB:	10
US-09-671-687A-3 (1-949) x BC042438 (1-4501)	
Qy	1 MetSerSerGlyLeuTrpSerGlnGluLysValThrSerProTyrTrpGluGluArgIle 20
Db	176 ATGAGTTCCAGGCTGTGGAGCCAGAGAAAGTTACTTCACTTCTGCGAAGCGGATT 235
Qy	21 PheTyrLeuLeuGluGluCysSerValThrAspLysGlnThrGlnLysLeuLys 40
Db	236 TTTTATCTGCTCTTCAGAAATGCAAGTGTAAACAGAAACAACTCAGAAAGCTGCTGAAA 295
Qy	41 ValProLysGlySerIleGlyGlnTyrIleGlnAspArgSerValGlyHisSerArgIle 60
Db	296 GTACCCAAAGGAGATAGGACAGTACATCCAAAGACCGTTCGTGGGCAATCAAGAGTT 355
Qy	61 ProSerAlaLysGlyLysLysAsnGlnIleGlyLeuLysIleLeuGluGlnProHisAla 80
Db	356 CTTTCCACAAAAGGCAAGAAAATCAGATTGGAATTAATAAATCTTGAGCAACCGCATGCA 415
Qy	81 ValLeuPheValAspGlu---AspValValGluIleAsnGluLysPheThrGluLeuLeu 99
Db	416 GTTCTGTTTGTGATAAAGGATGTTGTAGAAATAAATAAGAAATAATCAGAGATTACTG 475
Qy	100 LeuAlaIleThrAsnCysGluGluArgPheSerLeuPheLysAsnArgAsnArgLeuSer 119
Db	476 TTGGCAATTACCACTGTGAGGAGAGGCTCACCTTATTAGAAACAGACTCCGACTAAGT 535
Qy	120 LysGlyLeuGlnIleAspValGlyCysProValLysValGlnLeuArgSerGlyGluGlu 139
Db	536 AAAGGCTCCAGCTAGACGTGGGCAGTCTCTGTGAAAGTACAGCTGCGATCTGGGGAAGAG 595
Qy	140 LysPheProGlyValValArgPheArgGlyProLeuLeuAlaGluArgThrValSerGly 159

```
Db 1673 CTGGAACGGAAGATGAATCGGACGGCTGTACAGATGGAACCTTCAGGGGACCGCGTAT 1732
Qy 517 PheThrCysAlaLeuLysLysAlaLeuPheValLysLeuLysSerCysArgProAspSer 536
Db 1733 TTCACGTGTGCCCTGAAGAGGCACCTGTTGTGAAACTGAAGAGCTGCAGACCGGACTCT 1792
Qy 537 ArgPheAlaSerLeuGlnProValSerAsnGlnIleGluArgCysAsnSerLeuAlaPhe 556
Db 1793 AGGTGTCATCCTTGCAGCCTGTTTCCAATCAAAATTGAAAGGTGTAACTCTTTAGCATTT 1852
Qy 557 GlyGlyTyrLeuSerGluValValGluGluAsnThrProProLysMetGluLysGluGly 576
Db 1853 GGGGGCTATTAAAGTAGTAGTAGAAGAAATACTCCACTAAATGGAAGAAAGAGGT 1912
Qy 577 LeuGluIleMetIleGlyLysLysLysGlyIleGlnGlyHisTyrAsnSerCysTyrLeu 596
Db 1913 TTAGAGATAATGATTGGAAAGAGAAGAGGCATCCAGGGCCATTACAACTTCTGTACTTA 1972
Qy 597 AspSerThrLeuPheCysLeuPheAlaPheSerSerValLeuAspThrValLeuLeuArg 616
Db 1973 GACTCAACTTATTCTGCTTATTGCTTTAGTCTGCTCCCTGGACACTGTGTACTTAGA 2032
Qy 617 ProLysGluLysAsnAspValGluTyrTyrSerGluThrGlnGluLeuLeuArgThrGlu 636
Db 2033 CCCAAAGAGAAGAAATATATAGAGTATTACAGTGAGACTCAGGAGCTACTGAAGACAGAG 2092
Qy 637 IleValAsnProLeuArgIleTyrGlyTyrValCysAlaThrLysIleMetLysLeuArg 656
Db 2093 ATAGTCAATCCTCTGAGATATATGATATGATGTGTGCCACAAGATTATGAACTGAGG 2152
Qy 657 LysIleLeuGluLysValGluAlaSerGlyPheThrSerGluGluLysAspProGlu 676
Db 2153 AAAATCTTTGAAAGTTGAGGCTGCATCAGGTTTTACCTCTGAAGAAAAAGATCTCTGAA 2212
Qy 677 GluPheLeuAsnIleLeuPheHisIleLeuArgValGluProLeuLeuLysIleArg 696
Db 2213 GAATCTTAAATATCTGTTTTCATGATATTTAAGGGTTGAACCAATGTTAAAAATAAGA 2272
Qy 697 SerAlaGlyGlnLysValGlnAspCysTyrPheTyrGlnIlePheMetGluLysAsnGlu 716
Db 2273 TCAGCAGGTCAAAAGTTCAAGACTGTAACTTCTATCAAAATTTTATGGAAAAAATATGAG 2332
Qy 717 LysValGlyValProThrIleGlnGlnLeuLeuGluTyrPheIleAsnSerAsnLeu 736
Db 2333 AAAAGTTGAGTACCCACAACTTACAGAGTATTAGATGCTGCTTTTATCAACAGCACTG 2392
Qy 737 LysPheAlaGluAlaProSerCysLeuIleIleGlnMetProArgPheGlyLysAspPhe 756
Db 2393 AAAATTTGCAGAGGCACCATCATGCTTGATTATCCAGATGCTCGGTTTGGAAAAAGACTTT 2452
Qy 757 LysLeuPheIleLysIlePheProSerLeuGluLeuAsnIleThrAspLeuLeuGluAsp 776
Db 2453 AAACTATTAAAAAAATTTTCTCTCCCTGGAATTAATAATAACAGATTTACTTGAAGAC 2512
Qy 777 ThrProArgGlnCysArgIleCysGlyGlyLeuAlaMetTyrGluCysArgGluCysTyr 796
Db 2513 ACTCCAGCAGGATGCCGCATCTCTGGAGGACTTGCAGATGACAGTGCCAGGAGTGCTAT 2572
Qy 797 AspAspProAspIleSerAlaGlyLysIleLysGlnPheCysLysThrCysAsnThrGln 816
Db 2573 GACGATCCGGACATCTCAGCTGGGAAGATCAAGCAGTTCTGTGAAGACCTGCAGCACTCAG 2632
Qy 817 ValHisLeuHisProLysArgLeuAsnHisLysTyrAsnProValSerLeuProLysAsp 836
Db 2633 GTTCACTTTCATCCAGGAGGTTGAATCATTTCTATCATCCAGTATCATCTCCCAAGAC 2692
Qy 837 LeuProAspThrAspThrArgHisGlyCysIleProCysGlnAsnMetGluLeuPheAla 856
Db 2693 TTGCTGACTGGGACTGGAGACACGGCTGCATCCCTGTCAGAAAGATGGAGTTATTGTCT 2752
Qy 857 ValLeuCysIleGluThrSerHisTyrValAlaPheValLysTyrGlyLysAspAspSer 876
Db 2753 GTTCTCTGATAGAAACGACCACTATGTTGCTTTTGTGAGTATGGGAAGATGACTCT 2812
```

```
Qy 877 AlaTrpLeuPhePheAspSerMetAlaAspArgAspGlyGlyGlnAsnGlyPheAsnIle 896
Db 2813 GCCTGGCTTTTCTTTGACAGCATGCGGATCGAGATGCTGGTTCAGATGCTTCAACATT 2872
Qy 897 ProGlnValThrProCysProGluValGlyGluTyrLeuLysMetSerLeuGluAspLeu 916
Db 2873 CCACAAGTGACGCCCTGCCAGAAAGTGGAGAGTACTTTGAAGATGCTCTCTGGAGGACCTG 2932
Qy 917 HisSerLeuAspSerArgArgIleGlnGlyCysAlaArgLeuLeuCysAspAlaTyr 936
Db 2933 CACTCTTTGGATTCCAGAAGATTCAAGGCTGTGCGCGCAGACTTCTTTTGGATGCATAC 2992
Qy 937 MetCysMetTyrGlnSerProThrMetSerLeuTyrLys 949
Db 2993 ATGTGCAATGATACCAGAGTCCAAACCATGAGCCTGTACAAA 3031
```

```
RESULT 7
AX883937 2845 bp DNA linear PAT 17-DEC-2003
LOCUS Sequence 18842 from Patent EP1074617.
ACCESSION AX883937
VERSION AX883937.1 GI:40038838
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ihii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
TITLE Primers for synthesising full-length cDNA and their use
JOURNAL Patent: EP 1074617-A 18842 07-FEB-2001;
Research Association for Biotechnology (JP)
FEATURES
source Location/Qualifiers
1. 2845
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
452..2647
/notes="unnamed protein product"
/codon_start=1
/db_xref="GI:40038839"
/translation="MQVELPPELNSRVSLKVGETIBSGTIVIFCDVLPCKSLGYFVG
VMDNPIGNWDRFDGQVLCFACVETILLIHNDIIPESVTQERRPKLAFMRSGVG
DKGSSGNHKPKATGSDPNRNSSELFYTLNGSSVDSQPSQSKNTWYIDEVADPA
KSLTEISTDFDRSPPLQPPVNSLTTFENRPHSLPFLTKMPTNGSTIGHSPLSLQA
SVMEELNTAVQESPLAMPNGSHGVSLAEVKENPPFPYVIRWIGQPPGNEVL
AGLEDECACTDGTFRGTRYFTCALKKALFVKLKSRRPDRFASLQPPVNOIERCV
SLAFGYLSEVVEENTPPMKEGLEIMICKGIQGHYNSCYLDSILFCLFAFSSVL
DTVLPGKENDVEYSETOELLRTIENPLRIYGYCATKIMLKRKILEKVEASGF
TSBEKPEEFNLTFPHILREVPKLLRSAGQVQDCYFYQIFMEKNEKVGPVTIQOL
LEWSFINSLKFAEAPSLTIQMPRFQDKFLFKKIFPSLELNTDLELLETPROCIC
GGALVRECYDDPDISAGIKQPCFCTNTOVHLHPKRLNKNVPSLPKOLDPDWDW
RHGCI3CONMELEFVLCIETSHYVAVKYCKDSAWLFFDSMADRGGONGFNIPQVT
PCPEVEYGLKMSLEDLHLSRRIQGCARLLCLDAYMCMQSPHSLYK"
```

ORIGIN

```
Alignment Scores:
Pred. No.: 0 Length: 2845
Score: 4601.00 Matches: 875
Percent Similarity: 99.32% Conservative: 0
Best Local Similarity: 99.32% Mismatches: 2
Query Match: 91.40% Indels: 4
DB: 6 Gaps: 4
```

US-09-671-687A-3 (1-949) x AX883937 (1-2845)

```
Qy 73 LysIleLeuGluGlnProHisAlaValLeuPheValAspGlu---AspValValGluTle 91
Db 2 AAAATTCTAGAGCAACCTCATGCGATGCTCTCTTTCTGATGAAAAGGATGTTGTAGAGATA 61
```

QY 92 AsnGluLysPheThrGluLeuLeuAlaIleThrAsnCysGluGluArgPheSerLeu 111
DB 62 AATGAAAAGTTACAGAGTTACTTTTGGCAATTACCAATTGTGAGAGAGGTTACGCGTG 121
QY 112 PheLysAsnArgAsnArgLeuSerLysGlyLeuGlnIleAspValGlyCysProValLys 131
DB 122 TTTTAAAAACAGAAAACAGACTAAGTAAGTAAAGCCCTCCAAATAGACGTGGCGTGTCTGTGAAA 181
QY 132 ValGlnLeuLeuArgSerGlyGluGlyLysPheProGlyValValArgPheArgGlyProLeu 151
DB 182 GTACAGCTGAGATCTGGGGAAGAAAATTTCTGGAGTTGTACGTTTCAGAGAGCCCGTG 241
QY 152 LeuAlaGluArgThrValSerGlyIlePhePheGlyValGluLeuLeuGluGlyArg 171
DB 242 TTAGCAGAGAGACAGTCTCCGGAATATCTTTGGAGTTGAATTCCTGGAAGAAGTGTGT 301
QY 172 GlyGlnGlyPheThrAspGlyValTyrGlnGlyLysGlnLeuPheGlnCysAspGluAsp 191
DB 302 GGTCAAGGTTTCACTGACGGGTGTACCAAGGGAACAGCTTTTTCAGTGTGATGAAGAT 361
QY 192 CysGly----PheValAlaLeuAspLysLeuGluLeuIleGluAspAspThrAlaLeu 210
DB 362 TGTGGCGTGTGTTGTTCATTTGACAGCTAGAACTCATAGAAGATGATGACACTGCATTG 421
QY 211 GluSerAspTyrAlaGlyProGlyAspThrMetGlnValGluLeuProProLeuGluIle 230
DB 422 GAAAGTGATTACGAGGTTCTGGGACACAAATGCAGGTGCAATCTCTCTTTGGAAATA 481
QY 231 AsnSerArgValSerLeuLysGlyGlyGluThrIleGluSerGlyThrValIlePheCys 250
DB 482 AACTCCAGAGTTCTTTTGAAGGTGGAGAAAACAATAGAACTCTGGAACAGTTATATTCGT 541
QY 251 AspValLeuProGlyLysGluSerLeuGlyTyrPheValGlyValAspMetAspAsnPro 270
DB 542 CATGTTTTGCCAGGAAAAAGAAAGCTTAGGATATTTTGTGGTGTGGACATGATGAACCCCT 601
QY 271 IleGlyAsnTrpAspGlyArgPheAspGlyVal---LeuCysSerPheAlaCysValGlu 289
DB 602 ATTGGCAACTGGGATGGAAGATTGATGGAGTGCAGCTTTGTAGTTTTTCGGTGTGTGAA 661
QY 290 SerThrIleLeuLeuHisIleAsnAspIleIleProGluSerValThrGlnGluArgArg 309
DB 662 AGTACAATTTCTATTGCACATCAATGATATCATCCAGAGAGTGTGACGAGGAAGGAGG 721
QY 310 ProProLysLeuAlaPheMetSerArgGlyValGlyAspLysGlySerSerHisAsn 329
DB 722 CCTCCCAACTTGCCTTTATGTCAAGAGGTGTGGGACAAAGGTTTCATCCAGTCATAAT 781
QY 330 LysProLysAlaThrGlySerThrSerAspProGlyAsnArg---ArgSerGluLeuPhe 348
DB 782 AAACCAAAAGGCTACAGGATCTACCTCAGACCCCTGGAAATAGAAACAGATCTGAATTTATTT 841
QY 349 TyrThrLeuAsnGlySerSerValAspSerGlnProGlnSerLysSerLysAsnThrTrp 368
DB 842 TATACCTTAATGGGTCTTCTGTGACTCACAAACCAATCCAAATCAAAAAATACATGG 901
QY 369 TyrIleAspGluValAlaGluAspProAlaLysSerLeuThrGluIleSerThrAspPhe 388
DB 902 TACATTGATGAAGTTGCAGAAGACCCCTGCAAAATCTCTTACAGAGATATCTACAGACTTT 961
QY 389 AspArgSerSerProProLeuGlnProProProValAsnSerLeuThrThrGluAsnArg 408
DB 962 GACCGTTCTTCCACCACCTCCAGCCCTCTCTGTGAACTCACTGACCAACCCGAGAACAGA 1021
QY 409 PheHisSerLeuProPheSerLeuThrLysMetProAsnThrAsnGlySerIleGlyHis 428
DB 1022 TTCCACTTTTACCATTAGTCTCACCAAGATGCCCAATACCNAATGGAAGTATTGGCCAC 1081
QY 429 SerProLeuSerLeuSerAlaGlnSerValMetGluGluLeuAsnThrAlaProValGln 448
DB 1082 AGTCCACTTCTCTGTGAGCCAGTCTGTAATGGAAGAGCTAAACACTGCACCCCGTCCAA 1141
QY 449 GluSerProProLeuAlaMetProProGlyAsnSerHisGlyLeuValGlySerLeu 468

DB 1142 GAGAGTCCACCCCTTGGCCATGCCTCTCTGGAACTCACATGGTCTAGAAAGTGGGCTCATTTG 1201
QY 469 AlaGluValLysGlyAsnProPheTyrGlyValIleArgTrpIleGlyGlnProPro 488
DB 1202 GCTGAAGTTAAGGAGAACCCCTCTTTTCTATGGGGTAATCCGTGGATCGGTGAGCCACCA 1261
QY 489 GlyLeuAsnGluValLeuAlaGlyLeuGluLeuGluAspGluCysAlaGlyCysThrAsp 508
DB 1262 GGACTGAATGAAGTCTCGCTGGACTGGAACTGGGAAGATGAGTGTGCAGGCTGTACGGAT 1321
QY 509 GlyThrPheArgGlyThrArgTyrPheThrCysAlaLeuLysLysAlaLeuPheValLys 528
DB 1322 SGAACCTTCAGAGGACCTCGGTATTTTCACTGTGCGCTGAAGAGCGCTGTTGTGAAA 1381
QY 529 LeuLysSerCysArgProAspSerArgPheAlaSerLeuGlnProValSerAsnGlnIle 548
DB 1382 CTGAAGAGCTGCAGGCCCTGACTCTAGGTTTGATCATTTGACGCCGGTTTCCAATCAGATT 1441
QY 549 GluArgCysAsnSerLeuAlaPheGlyGlyTyrLeuSerGluValValGluGluAsnThr 568
DB 1442 GAGCGCTGTAACTCTTTAGCATTTGGAGGCTACTTAAAGTGAAGTAGTAGAAGAAATACT 1501
QY 569 ProProLysMetGluLysGluGlyLeuGluIleMetIleGlyLysLysLysGlyIleGln 588
DB 1502 CCACCACAAATGGAAAAGAGGCTTGGAGATAATGATTGGGAGAGAAAGGCAATCCAG 1561
QY 589 GlyHisTyrAsnSerCysTyrLeuAspSerThrLeuPheCysLeuPheAlaPheSerSer 608
DB 1562 GGTCAATTACAATCTTGTACTTAGACTCAACCTTATTCTGCTTATTTGCTTTTAGTTCT 1621
QY 609 ValLeuAspThrValLeuLeuArgProLysGluLysAsnAspValGluTyrTyrSerGlu 628
DB 1622 GTTCTGGACACTGTGTACTTGGACCCAAAGAAAAGAACGATGTAGAAATATTATAGTGA 1681
QY 629 ThrGlnGluLeuLeuArgThrGluIleValAsnProLeuArgIleTyrGlyTyrValCys 648
DB 1682 ACCCAGAGCTACTGAGGACAGAAATTTGTAATCTCTGAGAAATATATGATATGTGTGT 1741
QY 649 AlaThrLysIleMetLysLeuArgLysIleLeuGluLysValGluAlaAlaSerGlyPhe 668
DB 1742 GCCACAAAAATATGAAACTGAGGAAAAATACTTTGAAAAGGTGGAGGCTGCATCAGGATTT 1801
QY 669 ThrSerGluGluLysAspProGluGluPheLeuAsnIleLeuPheHisIleLeuArg 688
DB 1802 ACCTCTGAAGAAAAAGATCCTGAGGAATTTCTTGAATATTCTGTCTTCATCATATTTTAAAG 1861
QY 689 ValGluProLeuLeuLysIleArgSerAlaGlyGlnLysValGlnAspCysTyrPheTyr 708
DB 1862 GTAGAACCTTTTCTTAAAAATAAGATCAGCAGGTCAAAAGGTACAGATTTGTACTTCTAT 1921
QY 709 GlnIlePheMetGluLysAsnGluLysValGlyValProThrIleGlnGlnLeuLeuGlu 728
DB 1922 CAAATTTTATGAAAAAAATCAGAAAGTTGGCGTTTCCCACAATTCAGCAGTTGTTAGAA 1981
QY 729 TrpSerPheIleAsnSerAsnLeuLysPheAlaGluAlaProSerCysLeuIleIleGln 748
DB 1982 TGGTCTTTTATCAACAGTAACCTGAAATTTGAGAGGACCATCATGTCTGTGATTATTCAG 2041
QY 749 MetProArgPheGlyLysAspPheLysLeuPheLysLysIlePheProSerLeuGluLeu 768
DB 2042 ATGCGCTCGAATTTGGAAAAGACTTTAAACTATTTTAAAAAAATTTTTCTCTCTGGAATTA 2101
QY 769 AsnIleThrAspLeuLeuGluAspThrProArgGlnCysArgIleCysGlyGlyLeuAla 788
DB 2102 AATATAACAGATTTTACTTGAACACACTCCAGACAGTGCAGGATATGTGGAGGCTTGCA 2161
QY 789 MetTyrGluCysArgGluCysTyrAspAspProAspIleSerAlaGlyLysIleLysGln 808
DB 2162 ATGTATGAGTGTAGAGAATGCTACGACGATCGGACATCTCAGCTGGAAAAATCAAGCAG 2221
QY 809 PheCysLysThrCysAsnThrGlnValHisLeuHisProLysArgLeuAsnHisLysTyr 828

```

Db      2222 TTTTGTAAACCTGCAACACTCAAGTCCACCTTCATCCGAGAGGCTGAATCATATAATAT 2281
QY      829  AsnProValSerLeuProLysAspLeuProAspTyrArgHisGlyCysIlePro  848
Db      2282 AACCCAGTGTCACTTCCCAAGACTTACCCGACTGGGACTGGAGACAGCGCTGCATCCCT 2341
QY      849  CysGlnAsnMetGluLeuPheAlaValLeuCysIleGluThrSerHisTyrValAlaPhe  868
Db      2342 TGCAGAATAATGAGTTATTTGCTGTTCTCTGCATAGAAACAAGCCACTATGTTGCTTTT 2401
QY      869  ValLysTyrGlyLysAspSerAlaTyrPhePheAspSerMetAlaAspAtcAsp  888
Db      2402 GTCAAGTATGGGAAGACGATTCTGCTGCTCTCTTTGACAGCATGGCCGATCGGGAT 2461
QY      889  GlyGlyGlnAsnGlyPheAsnIleProGlnValThrProCysProGluValGlyGluTyr  908
Db      2462 GGTGGTCAGATGGCTTCAACATTCCTCAAGTACCCCATGCCAGAGATGAGNAGATAC 2521
QY      909  LeuLysMetSerLeuGluAspLeuHisSerLeuAspSerArgArgIleGlnGlyCysAla  928
Db      2522 TTGAAGACTGCTCTGGAAGACCTGCATTCTCTTGACTCCAGGAGAATCCAAGGCTGTGCA 2581
QY      929  ArgArgLeuLeuCysAspAlaTyrMetCysMetTyrGlnSerProThrMetSerLeuTyr  948
Db      2582 CGAAGACTGCTTTGTGATGCATATATGTGCATGTACCAGAGTCCAACAATGAGTTTGTAC 2641
QY      949  Lys 949
Db      2642 AAA 2644

RESULT 8
BD160617 2845 bp DNA linear PAT 17-JAN-2003
LOCUS     Primer for synthesizing full-length cDNA and use thereof.
DEFINITION
ACCESSION BD160617
VERSION    BD160617.1 GI:27866375
KEYWORDS   JP 2002191363-A/15460.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2845)
Ota.T., Isogai.T., Nishikawa.T., Hayashi.K., Saito.K., Yamamoto.J.,
Ishii.S., Sugiyama.T., Wakamatsu.A., Nagai.K. and Otsuki.T.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002191363-A 15460 09-JUL-2002;
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002191363-A/15460
PD 09-JUL-2002
PF 28-JUL-2000 JP 2000280990
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
PI KEIICHI NAGAI, TETSUJI OTSUKI
PC
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
10,
PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
(452)...(2644).
FT CDS
source
Location/Qualifiers
1..2845
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 2845
Score: 4601.00 Matches: 875
Percent Similarity: 99.32% Conservative: 0

```

```

Best Local Similarity: 99.32% Mismatches: 2
Query Match: 91.40% Indels: 4
DB: 6 Gaps: 4
US-09-671-687A-3 (1-949) x BD160617 (1-2845)

QY 73 LysIleLeuGluGlnProHisAlaValLeuPheValAspGlu---AspValValGluIle 91
Db 2 AAAATTCTAGAGCAACCTCATCGAGTTCTCTTTGTGATGAAAAGGATGTTGTAGAGATA 61
QY 92 AsnGluLysPheThrGluLeuLeuAlaIleThrAsnCysGluGluArgPheSerIleu 111
Db 62 AATGAAAAGTTTCAAGAGTTTACTTTTGGCAATTACCAATTGTGAGGAGGTTTCAGCCTG 121
QY 112 PheLysAsnArgAsnArgLeuSerLysGlyLeuGlnIleAspValGlyCysProValLys 131
Db 122 TTTAAAAACAGAAACACAGCTAAGTAAAGGCTCCAAATAGACGTGGGTCTCTCTGTGAAA 181
QY 132 ValGlnLeuArgSerGlyGluGluLysPheProGlyValValArgPheArgGlyProLeu 151
Db 182 GTACAGCTGAGATCTGGGGAAGAAAATTTCTCTGGAGTTGTACGCTTCAGAGGACCCCTG 241
QY 152 LeuAlaGluArgThrValSerGlyIlePhePheGlyValGluLeuLeuGluGluGlyArg 171
Db 242 TTAGCAGAGAGGACAGTCTCCGGAATATTTCTTTGGAGTTGAATTTGCTGGAAGAGTCT 301
QY 172 GlyGlnGlyPheThrAspGlyValTyrGlnGlyLysGlnLeuPheGlnCysAspGluAsp 191
Db 302 GGTCAAGGTTTCACTGACGGGGTGTACCAAGGGAACACAGCTTTTTCAGTGTGATGAAGAT 361
QY 192 CysGly---PheValAlaLeuAspLysLeuGluLeuIleGluAspAspThrAlaLeu 210
Db 362 TGTGGCGTGTGTTGTTGTTGGACAAGCTAGAACTCATAGAAGATGATGACACTGCAATG 421
QY 211 GluSerAspTyrAlaGlyProGlyAspThrMetGlnValGluLeuProProLeuGluIle 230
Db 422 GAAAGTGATTACGCAAGTCTCTGGGACACAAATGAGGTCGAACTTCTCTCTTTGGAAATA 481
QY 231 AsnSerArgValSerLeuLysGlyGluThrIleGluSerGlyThrValIlePheCys 250
Db 482 AACTCCAGAGTTTCTTTGAAAGGTTGGGAACAACAATAGAACTTGAACAGTTATATTCTGT 541
QY 251 AspValLeuProGlyLysGluSerLeuGlyTyrPheValGlyValAspMetAspAsnPro 270
Db 542 GATGTTTTCAGGAAAAGAAAGCTTAGGATATTTTGTGTGGCATGTGATAACCT 601
QY 271 IleGlyAsnTrpAspGlyArgPheAspGlyVal---LeuCysSerPheAlaCysValGlu 289
Db 602 ATTTGGCAACTGGGATGGAAGATTTGATGGAGTGCAGCTTTGTAGTTTTTGGCTGTGAA 661
QY 290 SerThrIleLeuLeuHisIleAsnAspIleIleProGluSerValThrGlnGluArg 309
Db 662 AGTACAAATTTATTTGCACATCAATGATATCATCCAGAGAGTGTGACGCAAGAGGAGG 721
QY 310 ProProLysLeuAlaPheMetSerArgGlyValGlyAspLysGlySerSerHisAsn 329
Db 722 CTCTCCCAAACTTCCTTTATGTCAAGAGGTGTGGGACAAAGGTTTCATCAGATCATAT 781
QY 330 LysProLysAlaThrGlySerThrSerAspProGlyAsnArg---ArgSerGluLeuPhe 348
Db 782 AAACCAAGGCTACAGGATCTACCTCAGACCTCGAANAATAGAAACAGATCTGAATTTAT 841
QY 349 TyrThrIleuAsnGlySerSerValAspSerGlnProGlnSerLysSerLysAsnThrTrp 368
Db 842 TATACCTTAAATGGGTCTTCTGTTGACTCAACAACCAATCCAAATCAAAAATACATGG 901
QY 369 TyrIleAspGluValAlaGluAspProAlaLysSerLeuThrGluIleSerThrAspPhe 388
Db 902 TACATTGATGAAGTTGCAAGAGACCTCTGCAAAAATCTCTTACAGAGATATCTACAGACT 961
QY 389 AspArgSerSerProProLeuGlnProProProValAsnSerLeuThrThrGluAsnArg 408
Db 962 GACCGTTCTTACCACCACCTCCAGCCTCTCTCTGTGAACCTCCTCTGACCAACGAGACAGA 1021

```

QY 409 PheHisSerLeuProPheSerLeuThrLysMetProAsnThrAsnGlySerIleGlyHis 428
 DB 1022 TTCACACTCTTACCAATTCAGTCTCCACCAAGATGCCAATACCANTGGAAGTATTGGCCAC 1081
 QY 429 SerProLeuSerLeuSerAlaGlnSerValMetGluGluLeuAsnThrAlaProValGln 448
 DB 1082 AGTCCACATTCTCTGTGCAGCCCGAGTCTGTAATGGAAGAGCTAAACACATGCACCCCGTCCAA 1141
 QY 449 GluSerProProLeuAlaMetProProGlyAsnSerHisGlyLeuGluValGlySerLeu 468
 DB 1142 GAGAGTCCACCCCTGGCCATGCTCTCTGGAACTCACATGGTCTAGAAGTGGGCTCATTG 1201
 QY 469 AlaGluValLysGluAsnProProPheTyrGlyValIleArgTyrIleGlyGlnProPro 488
 DB 1202 GCTGAAGTTTAAAGAAACCCCTCTTCTTAAGGGTAAATCCGTGGATCGGTGAGCCACCA 1261
 QY 489 GlyLeuAsnGluValLeuAlaGlyLeuGluLeuGluAspGluCysAlaGlyCysThrAsp 508
 DB 1262 GGACTGAAATGAAGTCTCGCTGAGACTGGAACTGGAAGATGAGTGTGAGGCTGTACGGAT 1321
 QY 509 GlyThrPheArgGlyThrArgTyrPheThrCysAlaLeuLysLysAlaLeuPheValLys 528
 DB 1322 GGAACCTTCAGAGGCACCTCGGTATTTCACTGTGCCCTGAAGAAGCGCTGTTTGTGAA 1381
 QY 529 LeuLysSerCysArgProAspSerArgPheAlaSerLeuGlnProValSerAsnGlnIle 548
 DB 1382 CTGAAGAGCTGCAGGCCCTGACTTGAAGTTTGATCATTCAGCCCGGTTTCCAAATCAGATT 1441
 QY 549 GluArgCysAsnSerLeuAlaPheGlyGlyTyrLeuSerGluValValGluGluAsnThr 568
 DB 1442 GAGCCCTGTAACTCTTAGCATTTGAGGCTACTTAAAGTGAAGTAGTAGAAGAAATACT 1501
 QY 569 ProProLysMetGluLysGluGlyLeuGluIleMetIleGlyLysLysGlyIleGln 588
 DB 1502 CCACCAAAAAATGGAAGAGCGCTGGAGATAATGATTGGGAAGAAAGCGCATCCAG 1561
 QY 589 GlyHisTyrAsnSerCysTyrLeuAspSerThrLeuPheCysLeuPheAlaPheSerSer 608
 DB 1562 GGTCAATTACAATTTCTGTACTTAGACTCAACCTTATTCTGCTTATTGCTTTTAGTTCT 1621
 QY 609 ValLeuAspThrValLeuLeuArgProLysGluLysAsnAspValGluTyrTyrSerGlu 628
 DB 1622 GTTCTGGACACTGTGTACTTGGACCCAAAGAAAGAACGATGAGAAATATTATAGTGAA 1681
 QY 629 ThrGlnGluLeuLeuArgThrGluIleValAsnProLeuArgIleTyrGlyTyrValCys 648
 DB 1682 ACCCAAGAGCTACTGAGGACAGAAATTTGTTAATCCTCTCAGATATATGATGTGT 1741
 QY 649 AlaThrLysIleMetLysLeuArgLysIleLeuGluLysValGluAlaAlaSerGlyPhe 668
 DB 1742 GCCACAAAAATTTATGAAACTGAGGAAAAATCTGAAAAAGGTGAGGCTGCATCAGGATTT 1801
 QY 669 ThrSerGluLysAspProGluGluPheLeuAsnIleLeuPheHisIleLeuArg 688
 DB 1802 ACCTCTGAAGAAAAAGATCCTCAGGAATTTCTTGAATATTCTGTTCATCATATTTTAAAG 1861
 QY 689 ValGluProLeuLeuLysIleArgSerAlaGlyGlnLysValGlnAspCysTyrPheTyr 708
 DB 1862 GTAGAACCTTTCTTAAAAATAGATCAGCAGGTCAAAAGGTCAACAGATTGTACTTCTAT 1921
 QY 709 GlnIlePheMetGluLysAsnGluLysValGlyValProThrIleGlnLeuLeuGlu 728
 DB 1922 CAAATTTTATGGAAAAAATGAGAAAGTTGGCGTTTCCCAATTCAGCAGTTGTAGAA 1981
 QY 729 TrpSerPheIleAsnSerAsnLeuLysPheAlaGluAlaProSerCysLeuIleGln 748
 DB 1982 TGGTCTTTTATCAACAGTAGTAACCTGAAATTTGAGAGGCCACCATCATGTCTGATTATTCAG 2041
 QY 749 MetProArgPheGlyLysAspPheLysLeuPheLysLysIlePheProSerLeuLeuLeu 768
 DB 2042 ATGCCTCGAATTTGGAAAAAGACTTTAACTATTATTAATAAATTTTTCCTTCTCTGGAATTA 2101

QY 769 AsnIleThrAspLeuLeuGluAspThrProArgGlnCysArgIleCysGlyGlyLeuAla 788
 DB 2102 AATATAACAGATTACTTTGAAGACACTCCAGACAGTGGCGGATATGTGGGGCTTGCA 2161
 QY 789 MetTyrGluCysArgGluCysTyrAspAspProAspIleSerAlaGlyLysIleLysGln 808
 DB 2162 ATGTATGAGTGTAGAGAAATGCTACGACGATCGGACATCTCAGCTGGAAAAATCAAGCAG 2221
 QY 809 PheCysLysThrCysAsnThrGlnValHisLeuHisProLysArgLeuAsnHisLysTyr 828
 DB 2222 TTTTGTAAACCTTGCACACATCAAGTCCACCTTCATCCGAAGGCTGAATCATATAATAT 2281
 QY 829 AsnProValSerLeuProLysAspLeuProAspTyrAspTyrArgHisGlyCysIlePro 848
 DB 2282 AACCCAGTGTCACTTCCCAAGACTTACCCGACTGGGACTGGAGACACGGCTGCATCCCT 2341
 QY 849 CysGlnAsnMetGluLeuPheAlaValLeuCysIleGluThrSerHisTyrValAlaPhe 868
 DB 2342 TGCCAGAAATATGGAGTTATTTCGTCTCTGCATAGAAACAAGCCACTATGTGCTTTT 2401
 QY 869 ValLysTyrGlyLysAspAspSerAlaTrpLeuPhePheAspSerMetAlaAspArgAsp 888
 DB 2402 GTGAAGTAGTGGGAAGAGCAATTCGCTGGCTCTTCTTTGACAGCATGCGCGATCGGAT 2461
 QY 889 GlyGlyGlnAsnGlyPheAsnIleProGlnValThrProCysProGluValGlyGluTyr 908
 DB 2462 GGTGTGCAAGTGGCTTCAACATTCCTCAAGTCACCCCATGCCCAAGTAGAGAGTAC 2521
 QY 909 LeuLysMetSerLeuGluAspLeuHisSerLeuAspSerArgArgIleGlnGlyCysAla 928
 DB 2522 TTGAAGATGCTCTCGAGAGACCTGCAATTCCTTGGACTCCAGAGAAATCAAGGCTGTGCA 2581
 QY 929 ArgArgLeuLeuCysAspAlaTyrMetCysMetTyrGlnSerProThrMetSerLeuTyr 948
 DB 2582 CGAAGACTGCTTGTGTGTCATATATGTGATGTACCATGTACCAAGTCCAACAATGAGTTGTAC 2641
 QY 949 Lys 949
 DB 2642 AAA 2644
 RESULT 9
 AK024348
 LOCUS Homo sapiens cDNA FLJ14286 fis, clone PLACE1006129, highly similar
 DEFINITION to Homo sapiens HSPC057 mRNA.
 ACCESSION AK024348
 VERSION AK024348.1 GI:10436712
 KEYWORDS oligo capping; fis (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
 Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
 Wagatsuma,M., Hosokiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
 Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
 Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
 Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
 Nakamura,Y., Nagahara,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
 NEDO human cDNA sequencing project
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2845)
 AUTHORS Isogai,T. and Otsuki,T.
 TITLE Direct Submission
 JOURNAL Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
 Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of
 International Trade and Industry of Japan; cDNA full insert
 sequencing; Research Association for Biotechnology; cDNA library
 construction; 5'- & 3'-end one pass sequencing and clone selection;
 Helix Research Institute (supported by Japan Key Technology Center

etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.

FEATURES

source
1. .2845
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACE1006129"
/tissue_types="placenta"
/clone_lib="PLACE1"
/notes="cloning vector: pME18SFL3"

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 2845
Score: 4601.00 Matches: 875
Percent Similarity: 99.32% Conservative: 0
Best Local Similarity: 99.32% Mismatches: 2
Query Match: 91.40% Indels: 4
DB: 9 Gaps: 4

US-09-671-687A-3 (1-949) x AK024348 (1-2845)

QY 73 LysfLeuGluGlnProHisAlaValLeuPheValAspGlu---AspValValGluLe 91
DB 2 AAAAATCTAGAGAACCTCATGAGATTCTCTTTGTTGATGAAAGGATGTTGAGAGATA 61
QY 92 AsnGluLysPheThrGluLeuLeuAlaIleThrAsnCysGluGluAArgPheSerLeu 111
DB 62 ATGAAAAGTTTCACAGATTACTTTTGGCAATACCAATTTGTGAGGAGAGTTTCAGCCTG 121
QY 112 PheLysAsnArgAsnArgLeuSerLysGlyLeuGlnIleAspValGlyCysProValLys 131
DB 122 TTTAAAACAGAAACAGACTAAGTAAAGGCTCCNAATAGACGTGGGCTCTCTGTGAAA 181
QY 132 ValGlnLeuArgSerGlyGluGluLysPheProGlyValValArgPheArgGlyProLeu 151
DB 182 GTACAGCTGAGATCTGGGGAAGAAAATTTCTCTGGAGTTGTACGCTTCAGAGGACCCCTG 241
QY 152 LeuAlaGluArgThrValSerGlyIlePhePheGlyValGluLeuGluGlyArg 171
DB 242 TTAGCAGAGAGGACGTCTCCGGAATATCTTTGGAGTTGAATTTGCTGGGAAGAGTCGT 301
QY 172 GlyGlnGlyPheThrAspGlyValTyrgInGlyLysGlnLeuPheGlnCysAspGluAsp 191
DB 302 GGTCAAGTTTCTCTACCGGGGTGTACCAAGGGAACAGAGTTTTCAGTGTGATGAAGAT 361
QY 192 CysGly---PheValAlaLeuAspLysLeuGluLeuIleGluAspAspThrAlaLeu 210
DB 362 TGTGGCGTGTGTTGTCATTGGACAAGCTAGAACTCATAGAAGATGATGACACTGCATTG 421
QY 211 GluSerAspTyraAlaGlyProGlyAspThrMetGlnValGluLeuProLeuGluLe 230
DB 422 GAAAGTGATTACGCAGGTCTCGGGGACACAATGCAGGTGCAACTCTCTCTTTGGAAATA 481
QY 231 AsnSerArgValSerLeuLysGlyGluThrIleGluSerGlyThrValIlePheCys 250
DB 482 AACTCCAGAGTTTCTTTGAAGGTTGGAGAAACAATAGAACTCTGGAACAGTTATTTCTGT 541
QY 251 AspValLeuProGlyLysGluSerLeuGlyTyraPheValGlyValAspMetAspAsnPro 270
DB 542 GATGTTTTCAGGAAAAGAAAGCTTAGGATATTTTGGTGTGGACATGGATACCCCT 601
QY 271 IleGlyAsnTrpAspGlyArgPheAspGlyVal---LeuCysSerPheAlaCysValGlu 289
DB 602 ATTGGCAACTGGGATGGAAGATTGATGGAGTCAGCTTGTAGTTTGGCTGTGTTGAA 661
QY 290 SerThrIleLeuLeuHisIleAsnAspIleIleProGluSerValThrGlnGluAArg 309
DB 662 AGTACAAATTTCTATTGCACATCATATGATATCATCCACAGAGAGTGTACGCGAGGAAGAGG 721
QY 310 ProProLysLeuAlaPheMetSerArgGlyValGlyAspLysGlySerSerHisAsn 329

DB 722 CCTCCCAAACTTGCTTTATGTCAGAGGTGTGGGGACAAAGGTTTCATCCAGTCATAAT 781
QY LysProLysAlaThrGlySerThrSerAspProGlyAsnArg---ArgSerGluLeuPhe 348
DB 782 AAACCAAAAGGCTACAGGATCTACTCAGACCTCGAATATAGAAACAGATCTGAATATTT 841
QY TyrThrLeuAsnGlySerSerValAspSerGlnProGlnSerLysSerLysSerLysThrTrp 368
DB 842 TATACCTTAAATGGGTCTTCTGTTGACTCAACAACCAATCAAAATCAAAATAATCATGG 901
QY TyrlleAspGluValAlaGluAspProAlaLysSerLeuThrGluIleSerThrAspPhe 388
DB 902 TACATTGATGAAGTTGCAGAAAGCCCTGCAAAATCTCTTACAGAGATATCTACAGACTTT 961
QY AspArgSerSerProProLeuGlnProProValAsnSerLeuThrThrGluAsnArg 408
DB 962 GACCGTTCTTACCAACCACTCAGCCCTCTCTGTGAACTACTGACCAACCGAAGAACAGA 1021
QY PheHisSerLeuProPheSerLeuThrLysMetProAsnThrAsnGlySerIleGlyHis 428
DB 1022 TTCCACTCTTTTACCATTCAGTCTCAACAAGATGCCAATACCAATGGAAGTATTGGCCAC 1081
QY SerProLeuSerLeuSerAlaGlnSerValMetGluGluLeuAsnThrAlaProValGln 448
DB 1082 AGTCCACTTTCTCTGTGAGCCAGTCTGTAATGGAAGAGCTAAACACTGCACCCGTCCAA 1141
QY GluSerProProLeuAlaMetProProGlyAsnSerHisGlyLeuGluValGlySerLeu 468
DB 1142 GAGAGTCCACCTTGGCCATGCCCTCTGGGAACCTACATGCTGAGAGTGGGCTCATTTG 1201
QY AlaGluValLysGluAsnProProPheTyrglyValIleArgTrpIleGlyGlnProPro 488
DB 1202 GCTGAAGTTAAGGAGAACCTCTCTTTCTATGGGGTATCCGTGGATCGGTGAGCCACA 1261
QY GlyLeuAsnGluValLeuAlaGlyLeuGluLeuGluAspGluCysAlaGlyCysThrAsp 508
DB 1262 GGACTGAATGAAGTCTCGCTGGAGCTGGAACCTGGAAGATGAGTGTGAGGCTGTACGGAT 1321
QY GlyThrPheArgGlyThrArgTyrgPheThrCysAlaLeuLysLysAlaLeuPheValLys 528
DB 1322 GGAACCTTTCAGAGCACTCGGTATTTCACTGTCCTCGCTGGAAGAGGCGCTGTTGTGAAA 1381
QY LeuLysSerCysArgProAspSerArgPheAlaSerLeuGlnProValSerAsnGlnIle 548
DB 1382 CTGAAAGAGCTGCAGCCCTGACTCTAGGTTTGATCATTTGAGCCGCTTCCATCAGATT 1441
QY GluArgCysAsnSerLeuAlaPheGlyTyrgLeuSerGluValValGluGluAsnThr 568
DB 1442 GAGCGCTGTAACCTTTTAGCATTTGGAGGCTACTTAAGTGAAGTAGTAGAAGAAATACT 1501
QY ProProLysMetGluLysGluGlyLeuGluIleMetIleGlyLysLysLysGlyIleGln 588
DB 1502 CCACCAAAAATGGAAGAAGGCTTGGAGATAATGATTGGGAAGAAGAGGCAATCCAG 1561
QY GlyHisTyraAsnSerCysTyrgLeuAspSerThrLeuPheCysLeuPheAlaPheSerSer 608
DB 1562 GGTCAATTACAAATTTCTGTTACTTAGACTCAACCTTATTTCTCTTATTTCCTTTAGTTCT 1621
QY ValLeuAspThrValLeuLeuArgProLysGluLysAsnAspValGluTyrgTyrgSerGlu 628
DB 1622 GTTCTGGACACTGTTTACTTTGGACCCCAAGAAAGAACGATGTAGATAATATTAGTGAA 1681
QY ThrGlnGluLeuLeuArgThrGluIleValAsnProLeuArgIleTyrgTyrgValCys 648
DB 1682 ACCCAAGAGCTACTGAGGACAGAAAATTTGTAATCTCTGAGAATATATGATATGTGTGT 1741
QY AlaThrIleMetLysLeuArgLysIleLeuGluLysValGluAlaAlaSerGlyPhe 668
DB 1742 GCCACAAAAATATGAAACTGAGGAAAATATCTTGAAGAGGTGGAGGCTGCATAGGATTT 1801
QY ThrSerGluGluLysAspProGluGluPheLeuAsnIleLeuPheHisIleLeuArg 688
DB 1802 ACCTCGAAGAAAAGATCTCTGAGAAATTTCTTGAATATTTCTGTTTCATATATTTAAGG 1861

```
Qy 689 ValGluProLeuLeuLysIleArgSerAlaGlyGlnLysValGlnAspCysTyrPheTyr 708
Db 1862 GTAGAACCTTTGCTAAAATAATAGATCACAGAGTCAAAAGGTACAAAGATTGTACTTCTAT 1921
Qy 709 GlnIlePheMetGluLysAsnGlnLysValGlyValProThrIleGlnLeuLeuGlu 728
Db 1922 CAAATTTTATGGAAAAAATGAGAAAGTGGCGTGTCCCAATTCAGCAGTTGTAGAA 1981
Qy 729 TrpSerPheIleAsnSerAsnLeuLysPheAlaGluAlaProSerCysLeuIleleGln 748
Db 1982 TGGTCTTTTATCAACAGTAACCTGAAATTTGAGAGGACCACCATCATGTCTGATTATTCAG 2041
Qy 749 MetProArgPheGlyLysAspPheLysLeuPheLysLysIlePheProSerLeuLeuLeu 768
Db 2042 ATGCTCTGATTTGGAAGAAGACTTTAAATATTTAAAAAAATTTTCTCTCTGGAATTA 2101
Qy 769 AsnIleThrAspLeuLeuGluAspThrProArgGlnCysArgIleCysGlyLeuAla 788
Db 2102 AATAACAGATTTACTTTGAAGACACTCCAGACAGTCCCGATATGTGGAGGCTTGCA 2161
Qy 789 MetTyrGluCysArgGluCysTyrAspAspProAspIleSerAlaGlyLysIleLysGln 808
Db 2162 ATGTATGAGTGTAGAGAATGCTACGACGATCCGACATCTCAGCTCGAAGAAATCAAGCAG 2221
Qy 809 PheCysLysThrCysAsnThrGlnValHisLeuHisProLysArgLeuAsnHisLysTyr 828
Db 2222 TTTTGTAAACCTGCAACACTCAAGTCCACCTTCATCCGAAGAGCTGAATCATAAATAT 2281
Qy 829 AsnProValSerLeuProLysAspLeuProAspTTPAspTTPArgHisGlyCysIlePro 848
Db 2282 AACCCAGTGTACTTCCCAAGACTTACCGACTGGGACTGGAGACACGGCTGCATCCCT 2341
Qy 849 CysGlnAsnMetGluLeuPheAlaValLeuCysIleGluThrSerHisTyrValAlaPhe 868
Db 2342 TGCCAGATATGGAGTTATTGCTGCTCTCTCATAGAAACAAGCACCATATGTTGCTTTT 2401
Qy 869 ValLysTyrGlyLysAspAspSerIleAlaThrPhePheAspSerMetAlaAspArgAsp 888
Db 2402 GTGAAGTATGGGAAGGAGCATTTCTCCCTGGCTCTTTCTTTGACAGCATGGCCCATCGGAT 2461
Qy 889 GlyGlyGlnAsnGlyPheAsnIleProGlnValThrProCysProGluValGlyGluTyr 908
Db 2462 GTGTGTGAGATGGCTTCAACATTTCTCAAGTCAACCCATGCCAGATGAGAGATGATAC 2521
Qy 909 LeuLysMetSerLeuGluAspLeuHisSerLeuAspSerArgArgIleGlnGlyCysAla 928
Db 2522 TTGAAGATGCTCTCGAAGACCTGCATTCCTTGGACTCCAGGAGAAATCCAGGCTGTGCA 2581
Qy 929 ArgArgLeuLeuCysAspAlaTyrMetCysMetTyrGlnSerProThrMetSerLeuTyr 948
Db 2582 CGAAGACTGCTTTGTGATGCATATATGTGCATGTATGTCATGTACAGAGTCCAACATGATGTTGTAC 2641
Qy 949 Lys 949
Db 2642 AAA 2644

RESULT 10
AC123449
LOCUS
DEFINITION Rattus norvegicus clone CH230-21D6, *** SEQUENCING IN PROGRESS ***,
3 unordered pieces.
ACCESSION AC123449
VERSION AC123449.3 GI:23265406
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 241990)
Muzny,D,Marie., Metzker,M.Lee., Abranzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Aisbrooks,S., Amin,A., Anguitano,D.,
```

Anvalebachi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chaves, D., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gregoire, B., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, I., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuewa, L., Louised, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokelimeh, O., Okwuonu, G., Olanpunaagoo, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Popper, F., Poinexter, A., Popovic, D., Primus, E., Pu, L., Puazo, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Sherty, J., Shvartbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

2 (bases 1 to 241990)
Unpublished
Worley, K.C.
Direct Submission
Submitted (29-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 241990)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (21-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Sep 21, 2002 this sequence version replaced gi:21902787.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>


```

Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GXOF
Center clone name: CH230-21D6
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 229402 bases at least Q40
Consensus quality: 231763 bases at least Q30
Consensus quality: 233057 bases at least Q20
Estimated insert size: 260790; sum-of-contigs estimation
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
  (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
  * consists of 3 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
  * 1 238985: contig of 238985 bp in length
  * 238986 239085: gap of unknown length
  * 239086 240731: contig of 1646 bp in length
  * 240732 240831: gap of unknown length
  * 240832 241990: contig of 1159 bp in length.
  Location/Qualifiers
    source          1..241990
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-21D6"
    misc_feature    1323..2205
                    /note="clone_boundary"
                    clone_end:Sp6
                    site:EcoRI
                    end_sequence:BH269859"
    misc_feature    complement(236297..237124)
                    /note="clone_boundary"
                    clone_end:T7
                    site:EcoRI
                    end_sequence:BH269858"

ORIGIN
Alignment Scores:
Pred. No.:      0      Length:      241990
Score:          4600.50      Matches:      883
Percent Similarity: 95.29%      Conservative: 27
Best Local Similarity: 92.46%      Mismatches: 38
Query Match:      91.39%      Indels: 8
DB:              2      Gaps: 5

US-09-671-687A-3 (1-949) x AC123449 (1-241990)
QY      1 MetSerGlyLeuTrpSerGlnGluYsValThrSerProTyrTrpGluGluArgIle 20
Db      205375 ATGAGTTTCAGGCTGTGGAAACCAAGAGAAAGTTACTTCAACCTACTGGGAAGAACCGCTT 205434
QY      21 PheTyrLeuLeuGlnGluCysSerValThrAspLysGlnThrGlnLysLeuLeuLys 40
Db      205435 TTTTATCTGCTTCTCAAGATGCAATGCTGTAAAC-----AAACAGACACAGAAAGCTCTTGAGA 205490
QY      41 ValProLysGlySerIleGlyGlnTyrIleGlnAspArgSerValGlyHisSerArgIle 60
Db      205491 GTACCCAAAGGAGGACATAGACAGTACATCCAAAGACTGTTCCGTCGGGCATTCAGAGTT 205550
QY      61 ProSerAlaLysGlyLysAsnGlnIleGlyLeuLysIleLeuGlnProHisAla 80
Db      205551 CCTTCTGCTTAAAGGCAAGAAATCATGATTGGATTAAATAATCTTAGAGCAACCGCATGCA 205610
QY      81 ValLeuPheValAspGlu---AspValValGluIleAsnGluLysPheThrGluLeuLeu 99

```

```

Db      205611 GTTCTGTTTGTGATGAAAGGATGTTTGTAGAAATAAATGAAAAATTCACAGAGTTACTGT 205670
QY      100 LeuAlaIleThrAsnCysGluGluArgPheSerLeuPheLysAsnArgAsnArgLeuSer 119
Db      205671 TTGCAATATTCACCAACTGTGAGGAGAGGCTCAGCCCTATTATTAGAAACAGAAATCCGACTAAGT 205730
QY      120 LysGlyLeuGlnIleAspValGlyCysProValLysValGlnLeuArgSerGlyGluGlu 139
Db      205731 AAAGGCTCCAGGTAGAGTGGGCAGTCTCTGTGAGAGTACAGTCCGATCTGGGAGGAG 205790
QY      140 LysPheProGlyValValArgPheArgGlyProLeuLeuAlaGluArgThrValSerGly 159
Db      205791 AAGTTTCAGGAGTTGTACGCTTCAGAGGACCTTTATTAGCGAGAGGACCGGTGTCGGG 205850
QY      160 IlePhePheGlyValGluLeuLeuGluGluGlyArgGlyGlnGlyPheThrAspGlyVal 179
Db      205851 ATTTCTTTGGAGTAGAATTAATCGAAGAAAGTCTGTGCCAAGGTTTCACTGATGGGGT 205910
QY      180 TyrGlnGlyLysGlnLeuPheGlnCysAspGluAspCysGly---PheValAlaLeuAsp 198
Db      205911 TATCAAGGAAACACAGCTCTTCAGTGTGATCAGGACTGTGGTGTGTTGTTGTCATTTGAC 205970
QY      199 LysLeuGluLeuIleGluAspAspThrAlaLeuGluSerAspTyrAlaGlyProGly 218
Db      205971 AAGCTGGAGCTTATAGAAAGATGATGACAATGGGTGGAAAGTGAATTTTGCAGGCCAGGA 206030
QY      219 AspThrMetGlnValGluLeuProLeuGluIleAsnSerArgValSerLeuLysGly 238
Db      206031 GATACAGTCCAGGTTGAACCTCCCTTTGAAAATAAACTCCAGAGTTCTTTTGAAGTT 206090
QY      239 GlyGluThrIleGluSerGlyThrValIlePheCysAspValLeuProGlyLysGluSer 258
Db      206091 GGAGAAAGTACAGAAATCTGGAAACAGTGAATTTCTGTGATGTTTACCAGAAAAAGAGAT 206150
QY      259 LeuGlyTyrPheValGlyValAspMetAspAsnProIleGlyAsnTrpAspGlyArgPhe 278
Db      206151 CTAGGATATTTTGTGTGGTGGACATGATAACCTATTGGCAACTGGGATGGAAGTTT 206210
QY      279 AspGlyVal---LeuCysSerPheAlaCysValGluSerThrIleLeuLeuHisIleAsn 297
Db      206211 GATGGGGTTTCAGCGTTTCAGGTTTTGCAAGTGTTCAGAGTACAGTTCTCTTACACATCAAT 206270
QY      298 AspIleIleProGluSerValThrGlnGluArgArgProProLysLeuAlaPheMetSer 317
Db      206271 GACATCATCCAGATACGTCGACACAGAAAGGAGACCTCCCAAACTTGCCTTTATGTCA 206330
QY      318 ArgGlyValGlyAspLysGlySerSerHisAsnLysProLysAlaThrGlySerThr 337
Db      206331 AGAGGTGTAGGTGACAAAGGCTCATCTAGTCATATAAACCAAGGTTACAGGATCTACC 206390
QY      338 SerAspProGlyAsnArg---ArgSerGluLeuPheTyrThrLeuAsnGlySerSerVal 356
Db      206391 TCAGACCTCGAAGTAGAACAACAGACTCTGAATATTTTATATACCTTAAATGGGTCTCATCTGT 206450
QY      357 AspSerGlnProGlnSerLysSerLysAsnThrTrpTyrIleAspGluValAlaGluAsp 376
Db      206451 GACTCACAAACAATCCAAAGTCCAAAACCCATGGTACATTGATGAAGTTGCAGAAAGAC 206510
QY      377 ProAlaLysSerLeuThrGluIleSerThrAspPheAspArgSerSerProProLeuGln 396
Db      206511 CCTGCAAAAGTCATTACAGAGATGCTTCAGACTTCGGACATTCATCGCTCCCTCCCGCNA 206570
QY      397 ProProProValAsnSerLeuThrGluAsnArgPheHisSerLeuProPheSerLeu 416
Db      206571 CCTCTCTCCATGAATCCTTGTCTAGCGAGAACAGATTCCACTCTTACCCTTCCAGCCTG 206630
QY      417 ThrLysMetProAsnThrAsnGlySerIleGlyHisSerPro---LeuSerLeuSerAlaG 436
Db      206631 ACAAAAGATGCCAACCACTAATGGCAGCATGGCTCACAGTCCACTCTCTCTGTGAGTGC 206690
QY      436 lnSerValMetGluGluLeuAsnThrAlaProValGlnGluSerProProLeuAlaMetP 456

```

Db 206691 AGTCTGTGATGGGGAGCTGAACAGCAGCCCTGTCCAGGAGAGTCCACCCTTGGCCAGCT 206750
Qy 456 roProGlyAenSerHisGlyLeuGluValGlySerLeuAlaGluValLysGluAenProp 476
Db 206751 CTTCTGGGAATGGACACGGGCTAGAGGTGGGCTCACTGGCTGAAGTAAAGAGAACCCCC 206810
Qy 476 roPheTyrGlyValIleArgTnpIleGlyGlnProProGlyLeuAenGluValLeuAlaG 496
Db 206811 CGTTCATGGGGTATCCGTTGGATGGCCAGCCACAGGGCTCAGTGCAGCTGCTCGTG 206870
Qy 496 lyLeuGluLeuGluAspGluCysAlaGlyCysThrAspGlyThrPheArgGlyThrArgT 516
Db 206871 GATTGGAACCGGAACATGAATGCGCAGGTTCACGGATGGAATTTTCAGGGCACGCGCT 206930
Qy 516 yrPheThrCysAlaLeuLysLysAlaLeuPheValLysLeuLysSerCysArgProAspS 536
Db 206931 ATTTCCACCTGTGCCCTGAAGAAAGCACTGTTCTGTGAAACTGAAGAGCTGCACACAGACT 206990
Qy 536 erArgPheAlaSerLeuGlnProValSerAsnGlnIleGluArgCysAenSerLeuAlaP 556
Db 206991 CTAGGTTTGCACTCTTGAGCCCTGTTTCCAATCAGATGAAAGGTGTAACCTCTTTAGCAT 207050
Qy 556 heGlyGlyTyrLeuSerGluValValGluGluAsnThrProProLysMetGluLysGluG 576
Db 207051 TTGGGGGCTACTTAAGTGAAGTAGTAGAAGAAATACGCCACCTAAATGGAAAGGAAG 207110
Qy 576 lyLeuGluIleMetIleGlyLysLysLysGlyIleGlnGlyHisTyrAenSerCysTyrL 596
Db 207111 GTTTAGAGATATGATGGAAGAGAAAGAGGATCCAGGGCCATTAACAATTTCTGTACT 207170
Qy 596 euAspSerThrLeuPheCysLeuPheAlaPheSerSerValLeuAspThrValLeuA 616
Db 207171 TAGACTCAACTTATCTGCTTATTTGCTTTTAGTTCTGCTCGACACTGTATTACTTA 207230
Qy 616 rgProLysGluLysAsnAspValGluTyrTyrSerGluThrGlnGluLeuLeuArgThrG 636
Db 207231 GACCCAAAGAGAGAAATGACGTAGAGTATTACAGTGAGACTCAAGAGCTACTGAGGACAG 207290
Qy 636 luIleValAenProLeuArgIleTyrGlyTyrValCysAlaThrIysIleMetLysLeuA 656
Db 207291 AGATAGTCAATCTCTGAGAATATATGATATGTGTGTGCCCAAAAGATTATGAAGCTGA 207350
Qy 656 rgLysIleLeuGluLysValGluAlaAlaSerGlyPheThrSerGluGluLysAspProG 676
Db 207351 GGAATAATCTTGAAAAGTTGAGGTGCATCAGGATTTACCTCTGAGGAAAAGATCCCTG 207410
Qy 676 luGluPheLeuAenIleLeuPheHisIleLeuArgValGluProLeuLeuLysIleA 696
Db 207411 AAGAATTTCTAAACATCTCTGTTTCAATGATATTTTAAAGGATGACCATTTGTAAAAATAT 207470
Qy 696 rgSerAlaGlyGlnLysValGlnAspCysTyrPheTyrGlnIlePheMetGluLysAsnG 716
Db 207471 GGTCAAGCAGGTCAAAAAGTTCAAGACTGTAACCTTCTATCAAAATTTTATGGAATAAATG 207530
Qy 716 luLysValGlyValProThrIleGlnGlnLeuLeuGluTrpSerPheIleAenSerAsnL 736
Db 207531 AGAAAATCGGAGTACCCCAATCCAGCAGTTATTAGAAATGTTCTTTTATCAACAGCAACC 207590
Qy 736 euLysPheAlaGluAlaProSerCysLysLeuIleGlnMetProArgPheGlyLysAspP 756
Db 207591 TGAATTTTCGGAGGAGCCACCATCATGCTTGATTTATCCAGATGCTTGTTGGGAAGACT 207650
Qy 756 heLysLeuPheLysIlePheProSerLeuGluLeuAenIleThrAspLeuLeuGluA 776
Db 207651 TTAAACTATTTAAAAAATTTTTTCTCTCCCTGGAATTTAAATATGACAGATTTACTTGAAG 207710
Qy 776 spThrProArgGlnCysArgIleCysGlyLeuAlaMetTyrGluCysArgGluCysT 796
Db 207711 ACACCTCCAGGAGTCCCGCATCTGTGGAGGACTCCGCCATGTATGATGATGATGAGTGTG 207770
Qy 796 yrAspAspProAspIleSerAlaGlyLysIleLysGlnPheCysLysThrCysAsnThrG 816
Db 207771 ATGATGACCCGGACATCTCGGAGGGAAGATCAAGCAGTTCTGTAAAGACTCGCAGCATC 207830

Qy 816 lnValHisLeuHisProLysArgLeuAenHisLysTyrAenProValSerLeuProLysA 836
Db 207831 AGTTTCACCTTCATCCAGAGACTGAATCACTTACCATCCAGTATCACTTCCCAAG 207890
Qy 836 spLeuProAspTrpAspTrpArgHisGlyCysIleProCysGlnAenMetGluLeuPheA 856
Db 207891 ACTTCCCGCAGCTGGGACTGGAGACATGGCTGCATCCCATGTCAGAAAGATGGAGTTATTG 207950
Qy 856 laValLeuCysIleGluThrSerHisTyrValAlaPheValLysTyrGlyLysAspAspS 876
Db 207951 CTGTGCTCTGCATAGAAAACAGCCACTATGTTGCTTTTGTGAAGTACGGGAAGGATGACT 208010
Qy 876 exAlaTrpLeuPhePheAspSerMetAlaAspArgAspGlyGlyGlnAenGlyPheAsnI 896
Db 208011 CTGCTCTGCTCTCTTTGACAGCATGGCTGATCGAGATGGTGTGATGAGATGGCTTCAACA 208070
Qy 896 leProGlnValThrProCysProGluValGlyGluTyrLeuLysMetSerLeuGluAspL 916
Db 208071 TTCCACAAGTGACACCCCTGCCAGAAAGTAGGAGATGACTTGAAGATGTCTCTGAGGACC 208130
Qy 916 euHisSerLeuAspSerArgArgIleGlnGlyCysAlaArgArgLeuLeuCysAspAlaT 936
Db 208131 TGCACTCTTTGACTCCAGAGGATTCAGGCTGTGCGCGCAGACTTCTTTGCGATGCAT 208190
Qy 936 yrMetCysMetTyrGlnSerProThrMetSerLeuTyrLys 949
Db 208191 ACATGTGATGTACCAGAGTCCACCATGAGCTGTGTACAAA 208231
RESULT 11
AC098162/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-3014, *** SEQUENCING IN PROGRESS ***,
3 unordered pieces.
AC098162
AC098162.7 GI:30522215
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 251132)
AUTHORS Muzny, D.Marie., Metzker, M.Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gegeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, X., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jacob, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshewa, L., Loulseghe, H., Lozard, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapa, P., Martin, K., Martin, R., Meenen, E.,
Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, N., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,

Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokeneme, O., Okwuonu, G., Olarunpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

AUTHORS

JOURNAL

TITLE

AUTHORS

JOURNAL

TITLE

AUTHORS

JOURNAL

TITLE

AUTHORS

JOURNAL

COMMENT

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 251132
Score: 4600.50 Matches: 883
Percent Similarity: 95.29% Conservative: 27
Best Local Similarity: 92.46% Mismatches: 38
Query Match: 91.39% Indels: 8
DB: 2 Gaps: 5

US-09-671-687A-3 (1-949) x AC098162 (1-251132)

Qy 1 MesSer3GlyLeuTrpSerGlnGluLysValThrSerProTyrTrpGluArgIle 20
Db 163209 ATGAGTTTCAGGCTGTGGACCAAGAGAAAGTTACTTCCCTACTGGGAGAACGGCTT 163150
Qy 21 PheTyrLeuLeuGlnGluCysSerValThrAspLysGlnThrGlnLysLeuLys 40
Db 163149 TTTTATCTGCTTCTTCAGGAATGAGTGTAAC----AAACAGACACAGAAAGTCTCTGAGA 163094
Qy 41 ValProLysGlySerIleGlyGlnTyrIleGlnAspArgSerValGlyHisSerArgIle 60
Db 163093 GTACCCAAAGGAGCATTAGGACAGTACATCAAGACTGTTCCGTGGGCGCATTCAGAGTT 163034
Qy 61 ProSerAlaLysGlyLysLysAsnGlnIleGlyLeuLysIleLeuGluGlnProHisAla 80
Db 163033 CCTTCTGCTAAAGGCAAGAAAAATCAGATTGGATTAAAAATCTTAGAGCAACCCGATGCA 162974
Qy 81 ValLeuPheValAspGlu----AspValValGluLeuLeuLysPheThrGluLeuLeu 99
Db 162973 GTTCTGTTTGTGATGAAAAGGATGTTGTAGAAATAAATGAAAAATTCACAGAGTTACTG 162914
Qy 100 LeuAlaIleThrAsnCysGluGluArgPheSerLeuPheLysAsnAsnArgLeuSer 119
Db 162913 TTGCAATATACCAACTGTGAGGAGAGGCTAGCCCTATTAGAAACAGAAATTCGACTAGT 162854
Qy 120 LysGlyLeuGlnIleAspValGlyCysProValLysValGlnLeuArgSerGlyGluGlu 139
Db 162853 AAAGGCTCCAGGTAGACGTGGGCGAGTCTCTGTGAGAGTACAGTCCGATCTGGGAGGAG 162794
Qy 140 LysPheProGlyValValArgPheArgGlyProLeuLeuAlaGluArgThrValSerGly 159
Db 162793 AAGTTTCAGGAGTTGTACGCTTCAGAGGACCTTTTATTAGCGGAGAGACGCTGTCGGGG 162734
Qy 160 IlePhePheGlyValGluLeuLeuGluGluCysArgGlyGlnGlyPheThrAspGlyVal 179
Db 162733 ATTTTCTTTGGAGTAGAAATTTACTGGAAGAAGGTCTGGGCCCAAGGTTTTCACGTAGGGGTG 162674
Qy 180 TyrGlnGlyLysGlnLeuPheGlnCysAspGluAspCysGly---PheValAlaLeuAsp 198
Db 162673 TATCAGAGAAAACAGCTCTTCCAGTGTGATGAGAGACTGTGGTGTGTTTGTTCATTGGAC 162614

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

AUTHORS

JOURNAL

TITLE

AUTHORS

JOURNAL

TITLE

AUTHORS

JOURNAL

TITLE

AUTHORS

JOURNAL

COMMENT

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 251132
Score: 4600.50 Matches: 883
Percent Similarity: 95.29% Conservative: 27
Best Local Similarity: 92.46% Mismatches: 38
Query Match: 91.39% Indels: 8
DB: 2 Gaps: 5

US-09-671-687A-3 (1-949) x AC098162 (1-251132)

Qy 1 MesSer3GlyLeuTrpSerGlnGluLysValThrSerProTyrTrpGluArgIle 20
Db 163209 ATGAGTTTCAGGCTGTGGACCAAGAGAAAGTTACTTCCCTACTGGGAGAACGGCTT 163150
Qy 21 PheTyrLeuLeuGlnGluCysSerValThrAspLysGlnThrGlnLysLeuLys 40
Db 163149 TTTTATCTGCTTCTTCAGGAATGAGTGTAAC----AAACAGACACAGAAAGTCTCTGAGA 163094
Qy 41 ValProLysGlySerIleGlyGlnTyrIleGlnAspArgSerValGlyHisSerArgIle 60
Db 163093 GTACCCAAAGGAGCATTAGGACAGTACATCAAGACTGTTCCGTGGGCGCATTCAGAGTT 163034
Qy 61 ProSerAlaLysGlyLysLysAsnGlnIleGlyLeuLysIleLeuGluGlnProHisAla 80
Db 163033 CCTTCTGCTAAAGGCAAGAAAAATCAGATTGGATTAAAAATCTTAGAGCAACCCGATGCA 162974
Qy 81 ValLeuPheValAspGlu----AspValValGluLeuLeuLysPheThrGluLeuLeu 99
Db 162973 GTTCTGTTTGTGATGAAAAGGATGTTGTAGAAATAAATGAAAAATTCACAGAGTTACTG 162914
Qy 100 LeuAlaIleThrAsnCysGluGluArgPheSerLeuPheLysAsnAsnArgLeuSer 119
Db 162913 TTGCAATATACCAACTGTGAGGAGAGGCTAGCCCTATTAGAAACAGAAATTCGACTAGT 162854
Qy 120 LysGlyLeuGlnIleAspValGlyCysProValLysValGlnLeuArgSerGlyGluGlu 139
Db 162853 AAAGGCTCCAGGTAGACGTGGGCGAGTCTCTGTGAGAGTACAGTCCGATCTGGGAGGAG 162794
Qy 140 LysPheProGlyValValArgPheArgGlyProLeuLeuAlaGluArgThrValSerGly 159
Db 162793 AAGTTTCAGGAGTTGTACGCTTCAGAGGACCTTTTATTAGCGGAGAGACGCTGTCGGGG 162734
Qy 160 IlePhePheGlyValGluLeuLeuGluGluCysArgGlyGlnGlyPheThrAspGlyVal 179
Db 162733 ATTTTCTTTGGAGTAGAAATTTACTGGAAGAAGGTCTGGGCCCAAGGTTTTCACGTAGGGGTG 162674
Qy 180 TyrGlnGlyLysGlnLeuPheGlnCysAspGluAspCysGly---PheValAlaLeuAsp 198
Db 162673 TATCAGAGAAAACAGCTCTTCCAGTGTGATGAGAGACTGTGGTGTGTTTGTTCATTGGAC 162614

***** NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
***** NOTE: This sequence may represent more than one clone.
***** NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
***** This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GWML
Center clone name: CH230-3014

----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 241087 bases at least Q40
Consensus quality: 243305 bases at least Q30
Consensus quality: 244914 bases at least Q20
Estimated insert size: 256850; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

***** NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
***** NOTE: This sequence may represent more than one clone.
***** NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
***** This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

QY	199	LysLeuGluLeuIleGluAspAspThrAlaLeuGluSerAspTyrAlaGlyProGly	218	QY	556	heGlyGlyTyrLeuSerGluValValGluGluAsnThrProProLysMetGluLysGluG	576
Db	162613	AAGCTGGAGCTTATAGAAGATGATGACAAATGGTTGGAAAGTGATTTTTCAGGCCAGGA	162554	Db	161533	TTGGGGCTACTTAACTGAGTAGTAGAAGAAATACGCCACCTAAAATGGAAAGGAAG	161474
QY	219	AspThrMetGlnValGluLeuProProLeuGluIleAsnSerArgValSerLeuLysGly	238	QY	576	lyLeuGluIleMetIleGlyLysLysGlyIleGlnGlyHisTyrAsnSerCysTyrL	596
Db	162553	GATACAGTCCAGGTGGAACCTCCCTTTGGAAATAAATCCAGAGTTCTTTGAAGGTT	162494	Db	161473	GTTTAGAGATAATGATTGGAAAGAAAGGATCCAGGGCCATTACAATTTCTTGTTACT	161414
QY	239	GlyGluThrIleGluSerGlyThrValIlePheCysAspValLeuProGlyLysGluSer	258	QY	596	euAspSerThrLeuPheCysLeuPheAlaPheSerSerValLeuAspThrValLeuLeuA	616
Db	162493	GGAGAAAGATACAGAACTCGAAGACGATATCTCTGATGTTTACACGAGAAAGAGAGT	162434	Db	161413	TAGACTCAACTTATTTCTGCTTATTTGCTTTTAGTTCTGCCCTGGACACTGATTACTTA	161354
QY	259	LeuGlyTyrPheValGlyValAspMetAspAsnProIleGlyAsnTrpAspGlyArgPhe	278	QY	616	rgProLysGluLysAsnAspValGluTyrTyrSerGluThrGlnGluLeuLeuArgThrG	636
Db	162433	CTAGATATTTTGTGTGTGACATGATTAACCTATTGGCACTGGGATGGAAAGGTTT	162374	Db	161353	GACCCAAAGAGAGAAATGACGTAGAGTATTACAGTGAAGACTCAGAGCTACTGAGGACAG	161294
QY	279	AspGlyVal---LeuCysSerPheAlaCysValGluSerThrIleLeuLeuHisIleAsn	297	QY	636	luIleValAsnProLeuLeuArgIleTyrGlyTyrValCysAlaThrIleMetLysLeuA	656
Db	162373	GATGGGTTTCAGCGTTGCAGTTTTCGAAGTGTTCGAGGTACAGTTCTCTCACATCAAT	162314	Db	161293	AGATAGTCAATCCTCTGAGAAATATATGGATATGTGTGTGCCACAAAGATTATGAAGCTGA	161234
QY	298	AspIleIleProGluSerValThrGlnGluArgArgProProLysLeuAlaPheMetSer	317	QY	656	rgLysIleLeuGluLysValGluAlaAlaSerGlyPheThrSerGluGluLysAspProG	676
Db	162313	GACATCATCCAGATAGCGTGACACAGGAAAGAGACCTCCCAAACTTGCCTTTATGTCA	162254	Db	161233	GGAAATACTTTGAAAAAGTTGAGGCTGCATCAGGATTTTACCTCTGAGGAAAAAGATCCTG	161174
QY	318	ArgGlyValGlyAspLysGlySerSerHisAsnLysProLysLeuAlaThrGlySerThr	337	QY	676	luGluPheLeuAsnIleLeuPheHisIleLeuArgValGluProLeuLeuLysIleA	696
Db	162253	AGAGGTGAGGTGACAAAGGCTCATCTAGTCATAATAAACCAAGGTTACAGGATCTACC	162194	Db	161173	ANGAATTTCTAAACATCTCTGTTTCATGATATTTTAAAGGATTTGACCTTGTTTAAAAATAT	161114
QY	338	SerAspProGlyAsnArg---ArgSerGluLeuPheTyrThrLeuAsnGlySerVal	356	QY	696	rgSerAlaGlyGlnLysValGlnAspCysTyrPheTyrGlnIlePheMetGluLysAsnG	716
Db	162193	TCAGACCTCGAAGTAGAAACAGATCTGATATTTTATACCTTAAATGGGTCTCTGTT	162134	Db	161113	GGTCAGCAGGTCAAAAAGTTCAAGACTGTAACTTCTATCAAAATTTTATGGAAAAAATG	161054
QY	357	AspSerGlnProGlnSerLysSerLysAsnThrTrpTyrIleAspGluValAlaGluAsp	376	QY	716	luLysValGlyValProThrIleGlnGlnLeuLeuGluTrpSerPheIleAsnSerAsnL	736
Db	162133	GACTCACAAACAATCAAGTCCAAAAACCCTCGTACATTGATGAAGTTCGAGAAAGAC	162074	Db	161053	AGAAAGTCGGAGTACCCCAATCCAGCAGATTATTAAGATTTCTTTTATCAACAGCAACC	160994
QY	377	ProAlaLysSerLeuThrGluIleSerThrAspPheAspArgSerSerProLeuGln	396	QY	736	euLysPheAlaGluAlaProSerCysLeuIleIleGlnMetProArgPheGlyLysAspP	756
Db	162073	CCTGCAAAAGTCACTTACAGAGATGCTTCAGACTTCGAGCATTCATCGCCTCCACCGCAA	162014	Db	160993	TGAAATTTTGGGAGGAGCCATCATCTGTTGATTATCCAGATGCTTGGTTTGGAAAGACT	160934
QY	397	ProProProValAsnSerLeuThrGluAsnArgPheHisSerLeuProPheSerLeu	416	QY	756	heLysLeuPheLysLysIlePheProSerLeuGluLeuAsnIleThrAspLeuLeuGluA	776
Db	162013	CCTCCTTCAGAACTCTTGTCTAGCGAAGAACAGATTCCACTCTTACCTTTCAGCCTG	161954	Db	160933	TTAAACTATTTAAAAAATTTTCTCTCCGGAATTAATAATATGACAGATTACTTCTGAAG	160874
QY	417	ThrLysMetProAsnThrAsnGlySerIleGlyHisSerPro--LeuSerLeuSerAlaG	436	QY	776	spThrProArgGlnCysArgIleCysGlyLysLeuAlaMetTyrGluCysArgGluCysT	796
Db	161953	ACAAAGATGCCCAACACTAATGGCAGCATGGCTCACAGTCCACTCTCTCTGTCTGCTGC	161894	Db	160873	ACACTCCCAGGAGTCCCGCATCTGTGGAGGACTCCGCAATGATGATGATGATGATGATG	160814
QY	436	InSerValMetGluGluLeuAsnThrAlaProValGlnGluSerProProLeuAlaMetP	456	QY	796	yrAspAspProAspIleSerAlaGlyLysIleLysGlnPheCysLysThrCysAsnThrG	816
Db	161893	AGTCTGTATGGGGAGCTGAACAGCAGCGCTGTCCAGAGAGTCCACCTTTGCCCGAGCT	161834	Db	160813	ATGATGACCCGAGCATCTCGCAGGGAAGATCAAGCAGTTCTGTAGACTCTGCAGCACTC	160754
QY	456	roProGlyAsnSerHisGlyLeuGluValGlySerLeuAlaGluValLysGluAsnProp	476	QY	816	InValHisLeuHisProLysArgLeuAsnHisLysTyrAsnProValSerLeuProLysA	836
Db	161833	CTTCTGGGAATGGACACGGGCTAGAGTGGGCTCACTGGCTGAGTAAAGAGAACCCCC	161774	Db	160753	AGTTTCACTTTCATCCAGAGACTGAATCAACACTTACCATTCCAGTATCACTTCCCAAG	160694
QY	476	roPheTyrGlyValIleArgTrpIleGlyGlnProProGlyLeuAsnGluValLeuAlaG	496	QY	836	spLeuProAspTrpAspTrpArgHisGlyCysIleProCysGlnAsnMetGluLeuPheA	856
Db	161773	CGTTCTATGGGTTTATCCGTTGGATGGCCAGCCACCGAGGCTCAGTGACGTGCTGCTG	161714	Db	160693	ACTTGGCCGACTGGGACGTGGAGCATGGCTGCATCCCATGTCAGAAAGATGGAGTTATTG	160634
QY	496	lyLeuGluLeuGluAspGluCysAlaGlyCysThrAspGlyThrPheArgGlyThrArgT	516	QY	856	laValLeuCysIleGluThrSerHisTyrValAlaPheValLysTyrGlyLysAspAspS	876
Db	161713	GATTTGAAACCGGAAGATGAATCGCAGGTTGCACGGATGGAACTTTTACGGGGCAGCGCT	161654	Db	160633	CTGTGCTCTGCAATAGAAACACCGCACTATGTTGCTTTTGTGAAGTACGGAGAGGATGACT	160574
QY	516	yrPheThrCysAlaLeuLysLysAlaLeuPheValLysLeuLysSerCysArgProAspS	536	QY	876	erAlaTrpLeuPhePheAspSerMetAlaAspArgAspGlyGlyGlnAsnGlyPheAsnI	896
Db	161653	ATTTCACCTGTGCTTGAAGAAAGCACTGTTTCTGTAACCTGAAGACTGCAGACCCAGCT	161594	Db	160573	CTGCTCTGGCTCTTCTTTGACAGCATGGCTGATCGAGATGGTGTGCAATGGCTTCAACA	160514
QY	536	erArgPheAlaSerLeuGlnProValSerAsnGlnIleGluArgCysAsnSerLeuAlaP	556	QY	896	leProGlnValThrProCysProGluValGlyLysTyrLeuLysMetSerLeuGluAspL	916
Db	161593	CTAGGTTTGCATCTTTCAGCGCTTTCCAATCAGATCGAAAGGTGTAATCTTCTAGCAT	161534	Db	160513	TTCCCAAGTGCACCCCTCCAGAGAGTAGGAGTACTTGAAGATGTCTCTGGAGACC	160454
				QY	916	euHisSerLeuAspSerArgIleGlnGlyCysAlaArgArgLeuLeuCysAspAlaT	936


```
Db 1665 ACAGAAATGCTACGACGATCCGACATCTCAGCTGGAATAATCAAGCAGTTTTTGTAAACCC 1724
Qy 813 CysAsnThrGlnValHisLeuHisProLysArgLeuAsnHisLeuAsnProValSer 832
Db 1725 TCAACACTCAAGTCCACCTTCATCGAAGAGCGTGAATCATATAATATAACCAAGTGCTCA 1784
Qy 833 LeuProLysAspLeuProAspTrpAspTrpArgHisGlyCysIleProCysGlnAsnMet 852
Db 1785 CTTCCCAAGACTTACCGACTGGGACTGGACACGCGTGCATCCCTTGCAGATATG 1844
Qy 853 GluLeuPheAlaValLeuCysIleGluThrSerHisTyrValAlaPheValLysTyrGly 872
Db 1845 CAGTTATTGTGCTGCTCTGCATAGAAACAAGCCACTATGTTGCTTTGTGAAGTATGG 1904
Qy 873 LysAspAspSerAlaTrpLeuPhePheAspSerMetAlaAspArgAspGlyGlnAsn 892
Db 1905 AAGGACGATTCCTGCTGCTCTCTTTGACACATGGCGCATGGCGATGGTGGT 1964
Qy 893 GlyPheAsnIleProGlnValThrProCysProGluValGlyGluTyrLeuLysMetSer 912
Db 1965 GGCCTTCAACATTCCTCAAGTCAACCCATGCCAGAGTAGGAGAGTACTTGAAGATGCT 2024
Qy 913 LeuGluAspLeuHisSerLeuAspSerArgArgIleGlnGlyCysAlaArgArgLeuLeu 932
Db 2025 CTGGAAGACCTGCATTCCTTGGACTCCAGGAGATCCAAAGGCTGTGCACGAAGACTGCTT 2084
Qy 933 CysAspAlaTyrMetCysMetTyrGlnSerProThrMetSerLeuTyrLys 949
Db 2085 TGTGATGCATATATGTGCATGTACCAGAGTCCAAACAATGAGTTGTACAAA 2135

RESULT 13
AK000187 2341 bp mRNA linear PRI 13-SEP-2003
LOCUS Homo sapiens cDNA FLJ20180 fis, clone COL10238, highly similar to
DEFINITION AB020656 Homo sapiens mRNA for KIAA0849 protein.
ACCESSION AK000187
VERSION AK000187.1 GI:7020107
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2341)
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1,
Minato-ku, Tokyo 108-8639, Japan (E-mail:flcdna@ims.u-tokyo.ac.jp,
Tel:81-3-5449-5286, Fax:81-3-5449-5416)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan: cDNA full insert
sequencing; Research Association for Biotechnology: cDNA library
construction; 5'- & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
Location/Qualifiers
1..2341
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="COL10238"
/tissue_type="colon"
/clone_lib="COL"
/note="cloning vector pME18SFL3"
```

```
misc_feature 1..2341
/note="highly similar to AB020656 Homo sapiens mRNA for
KIAA0849 protein"

ORIGIN
Alignment Scores:
Pred. No.: 2,32e-258 Length: 2341
Score: 3433.50 Matches: 644
Percent Similarity: 99.69% Conservative: 0
Best Local Similarity: 99.69% Mismatches: 1
Query Match: 68.21% Indels: 1
DB: 9 Gaps: 1
US-09-671-687A-3 (1-949) x AK000187 (1-2341)

Qy 305 ThrGlnGluArgArgProProLysLeuAlaPheMetSerArgGlyValGlyAspLysGly 324
Db 2 ACGCAGGAAGAGAGCGCTCCCAAACTTGCTTTATGTCAAGAGGTGTTGGGACAAAGGT 61
Qy 325 SerSerSerHisAsnLysProLysAlaThrGlySerThrSerAspProGlyAsnArg--- 343
Db 62 TCATCCAGTCATAATAAACCAAGGCTACAGGATCTACCTCAGACCCCTGGAATAGAAAC 121
Qy 344 ArgSerGluLeuPheTyrThrLeuAsnGlySerSerValAspSerGlnProGlnSerLys 363
Db 122 AGATCTGAATTATTTATACCTTAATGGGTCCTTCTGTGTGACTCACCAACCAATCCAA 181
Qy 364 SerLysAsnThrTrpTyrIleAspGluValAlaGluAspProAlaLysSerLeuThrGlu 383
Db 182 TCAAAAAATACATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 241
Qy 384 IleSerThrAspPheAspArgSerProProLeuGlnProProProValAsnSerLeu 403
Db 242 ATATCTACAGACTTTTGACCGCTTCTTCCACACACACCTCCAGCCCTCCTCTGTGAAC 301
Qy 404 ThrThrGluAsnArgPheHisSerLeuProPheSerLeuThrLysMetProAsnThrAsn 423
Db 302 ACCACCGAGACAGATTCCACTCTTTACCATTCAGTCTCACCAAGATGCCAATACCAAT 361
Qy 424 GlySerIleGlyHisSerProLeuSerLeuSerAlaGlnSerValMetGluGluLeuAsn 443
Db 362 GGAAGTATTGGCCACAGTCCACTTTCTCTGTGAGCCAGCTCTGTAATGGAAGAGCTAAAC 421
Qy 444 ThrAlaProValGlnGluSerProProLeuAlaMetProProGlyAsnSerHisGlyLeu 463
Db 422 ACTGCACCGCTCCCAAGAGAGTCCACCTTGGCCATGCCCTCTCCGGGAACCTCACAT 481
Qy 464 GluValGlySerLeuAlaGluValLysGluAsnProProPheTyrGlyValIleArgTyr 483
Db 482 GNAAGTGGGCTCATTTGGCTGAAGTTAAGGAGAACCCCTCTTCTATGGGGTAAATCCG 541
Qy 484 IleGlyGlnProProGlyLeuAsnGluValLeuAlaGlyLeuGluLeuGluAspGlyCys 503
Db 542 ATCGGTGACGCCACCAAGGACTGAATGAAGTGTCTGCTGGACTGGAACTGGAAGATGAT 601
Qy 504 AlaGlyCysThrAspGlyThrPheArgGlyThrArgTyrPheThrCysAlaLeuLysLys 523
Db 602 GCAGGCTGTACGGATGGAACCTTTCAGAGGCACTCGGTATTTCCACCTGTGCCCTGAAG 661
Qy 524 AlaLeuPheValLysLeuLysSerCysArgProAspSerArgPheAlaSerLeuGlnPro 543
Db 662 GGCCTGTTTGTGAACCTGAAGAGCTGCAGGCTGACTCTAGGTGTTGATTCATTGACGCC 721
Qy 544 ValSerAsnGlnIleGluArgCysAsnSerLeuAlaPheGlyGlyTyrLeuSerGluVal 563
Db 722 GTTTCCAATCAGATTGAGCGCTGTAACCTTTTAGCATTTGGAGGCTACTTAAAGTGAAG 781
Qy 564 ValGluGluAsnThrProProLysMetGluLysGluGlyLeuGluLeuMetIleGlyLys 583
Db 782 GTAGAAGAAAAATACCTCCCAAAAAATGGAAGAGGCTTGGAGATATATGATTGGGAAG 841
Qy 584 LysLysGlyIleGlnGlyHisTyrAsnSerCysTyrLeuAspSerThrLeuPheCysLeu 603
```

```
Db 842 AAGAAAGGCATCCAGGTCATTACAAATCTTGTTACTTACACTCAACCTTATCTGCTTA 901
Qy 604 PheAlaPheSerValLeuAspThrValLeuLeuArgProLysGluLysAsnAspVal 623
Db 902 TTTGCTTTTAGTTCTGTTCTGGACACTGTGTACTTAGACCCAAAGAAAGACGATGTA 961
Qy 624 GluTyrTyrSerGluThrGlnGluLeuLeuArgThrGluLeuValLeuAsnProLeuArgIle 643
Db 962 GAATATTATTAGTAAACCCAGAGCTACTGAGGACAGAAATTTGTTAATCTCTGAGAATA 1021
Qy 644 TyrGlyTyrValCysAlaThrLysIleMetLysLeuArgLysIleLeuGluLysValGlu 663
Db 1022 TATGGATATGTGTGTGCCACAAAAATATGAACCTGAGGAAATACTTGAAGAGTGGAG 1081
Qy 664 AlaAlaSerGlyPheThrSerGluGluLysAspProGluGluPheLeuAsnIleLeuPhe 683
Db 1082 GCTGCATCAGATTGCTCTGAAGAAAGATCTCTGAGGAATCTTGAATATCTGTTT 1141
Qy 684 HisHisIleLeuArgValGluProLeuLeuLysIleArgSerAlaGlyGlnLysValGln 703
Db 1142 CATCATATTTTAAGGGTAGAACCTTTGCTAAAAATAAGATCAGCAGGTCAAAAGGTACAA 1201
Qy 704 AspCysTyrPheTyrGlnIlePheMetGluLysAsnGluLysValGlyValProThrIle 723
Db 1202 GATTGTTACTTCTATCAAAATTTTATGAAAAAATGAGAAAGTTGGCGTTCCACAAAT 1261
Qy 724 GlnGlnLeuLeuGluTyrSerPheIleAsnSerAsnLeuLysPheAlaGluAlaProSer 743
Db 1262 CAGCAGTGTGTAGATGGTCTTTATCAACAGTAACCTGAAATTTGCAGAGGCCACCATCA 1321
Qy 744 CysLeuIleIleGlnMetProArgPheGlyLysAspPheLysLeuPheLysIlePhe 763
Db 1322 TGTCTGATTATTCAGATGCTCGATTGGAAGAAAGACTTTAAACTATTTAAAAAATTTT 1381
Qy 764 ProSerLeuLeuLeuAsnIleThrAspLeuLeuGluAspThrProArgGlnCysArgIle 783
Db 1382 CCTTCTCTGGAATTAATAATAACAGATTATTAATTGAAGACACTCCACAGACAGTCCGGATA 1441
Qy 784 CysGlyGlyLeuAlaMetTyrGluCysArgGluCysTyrAspAspProAspIleSerAla 803
Db 1442 TGTGGAGGGCTTCCAATGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1501
Qy 804 GlyLysIleLysGlnPheCysLysThrCysAsnThrGlnValHisLeuHisProLysArg 823
Db 1502 GGAATAATCAAGCAGTTTGTAAACCTTGCAACACTCAAGTCCACCTTCATCCGAGAGG 1561
Qy 824 LeuAsnHisLysTyrAsnProValSerLeuProLysAspLeuProAspThrAspTyrArg 843
Db 1562 CTGAATCATAAATAATAACCCAGTGTCACTTCCCAAGACTTACCCGACTGGGACTGGAGA 1621
Qy 844 HisGlyCysIleProCysGlnAsnMetGluLeuPheAlaValLeuCysIleGluThrSer 863
Db 1622 CAGCGTGTCACTCCCTTGGCAGATATGAGTATTTGCTGTTCTCTGCAATAGAAACAAGC 1681
Qy 864 HisTyrValAlaPheValLysTyrGlyLysAspAspSerAlaTyrLeuPhePheAspSer 883
Db 1682 CACTATGTTGCTTTTGTGAAGTATGGAAGAGAGATCTGCCCTGGCTCTCTTTTGACAGC 1741
Qy 884 MetAlaAspArgAspGlyGlnGlnAsnGlyPheAsnIleProGlnValThrProCysPro 903
Db 1742 ATGGCCGATCGGGATGGTGTGATGATGGTGTGATGGTGTGATGGTGTGATGGTGTGATGG 1801
Qy 904 GluValGlyGluTyrLeuLysMetSerLeuGluAspLeuHisSerLeuAspSerArgArg 923
Db 1802 GAAGTAGGAGAGTACTTGAAGATGTCTCTGGAAGAGAGCTGATTCCTTGGACTCCAGGAGA 1861
Qy 924 IleGlnGlyCysAlaArgArgLeuLeuCysAspAlaTyrMetCysMetTyrGlnSerPro 943
Db 1862 ATCCAAGGCTGTGCACGAAGACTGCTTTGTGTGATGATATATGATGATGATGATGATGATG 1921
Qy 944 ThrMetSerLeuTyrLys 949
Db 1922 ACATGAGTTTGTACAAA 1939
```

RESULT 14

```
AX883681 LOCUS 2569 bp DNA linear PAT 17-DEC-2003
Sequence 18586 from Patent EP1074617.
AX883681 DEFINITION
AX883681 ACCESSION
AX883681.1 VERSION AX883681.1 GI:40038582
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE Primers for synthesising full-length cDNA and their use
JOURNAL Patent: EP 1074617-A 18586 07-FEB-2001;
Research Association for Biotechnology (JP)
FEATURES
source
1. 2569
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
403..1833
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAE92065.1"
/db_xref="GI:40038583"
/translation="MQVELPPLEINSRVSLKVGETIESGTVIFCDVLPGKESLGYFVG
VMDNPIGNWDGRFQVQLCSFACVETILLHINDIIPESVTQRRPPKLAFMGRGVG
DKSSSHNPKPKATGSDTPGNRNRSLEFYTLNGSSVDSPQSKNTWYIDVADPDA
KSLTEITDTPDRSSPLQPPVNSLTATNENFSLTTPFSLTKMPNTNGSIGHSPLSAQ
SVMEELNTAPQESBPPLAMPNGSHGLEVSLAEVKENPPPYGVIRWIGQPPGLNEVL
AGLELDECACTDGTFRGTYRTCAIKKALFVKLSCDRSPFASQPVSNQLERCN
SLAFGYLSEVVEENTPPRMKEGLEIMIKKKGIQGHYNSCYLDSTLFCUFAFSSVL
DTLLRPKKNDEYVYSETQELLRTIVNPLRIYGVYCATKIMKLRLIKEVEASGF
TSBEKGDLNLVYALKITEIHMSKY"
ORIGIN
```

Alignment Scores:

```
Pred. No.: 3 52e-226 Length: 2569
Score: 3022.50 Matches: 587
Percent Similarity: 98.66% Conservative: 2
Best Local Similarity: 98.32% Mismatches: 3
Query Match: 60.04% Indels: 5
DB: 6 Gaps: 4
```

US-09-671-687A-3 (1-949) x AX883681 (1-2569)

```
Qy 88 ValValGluIleAsnGluLysPheThrGluLeuLeuAlaIleThrAsnCysGluGlu 107
Db 1 GTTGTAGAGATAAATAAGAAAGTTACACAGAGTTACTTTTGGCAATTACCAATTTGAGAG 60
Qy 108 ArgPheSerLeuPheLysAsnArgAsnArgLeuSerLysGlyLeuGlnIleAspValGly 127
Db 61 AGGTTACAGCTGTTTAAAAACACAGACAGACTAAGTAAAGCCCTCCAAATAGACGTGGGC 120
Qy 128 CysProValLysValGlnLeuArgSerGlyGluLysPheProGlyValValArgPhe 147
Db 121 TGTCTCTGTAAGATACAGCTGAGATCTGGGGAAGAAAAATTTCTCGAGTGTGACGCTTC 180
Qy 148 ArgGlyProLeuLeuAlaGluArgThrValSerGlyIlePhePheGlyValGluLeuLeu 167
Db 181 AGAGGACCCCTGTTAGCAGAGAGACAGACTCTCCGGAATATTTCTTGGAGTTGAATTCGTG 240
Qy 168 GluGluGlyArgGlyGlnGlyPheThrAspGlyValTyrGlnGlyLysGlnLeuPheGln 187
Db 241 GAAGAGGTCGTGTCAGGTTTTCAGCTGCGGGGTGACCAGGGAACACACCTTTTTCAG 300
Qy 188 CysAspGluAspCysGly----PheValAlaLeuAspLysLeuGluLeuIleGluAspAsp 206
Db 301 TGTGATGAAGATTGTGGCGTGTGTTGTTGTCATTGACAGACTAGAACTCATAGAGATGAT 360
```


Qy	207	AspThrAlaLeuGluSerAspTyrAlaGlyProGlyAspThrMetGlnValGluLeuPro	226
Db	361	GACACTGCATTGGAAAGTGATTCACGACGCTCCTGGGACACAATATCGCAGGTCCGAATTCCT	420
Qy	227	ProLeuGluIleAsnSerArgValSerLeuLysGlyGlyGluThrIleGluSerGlyThr	246
Db	421	CCTTTGGAAAATAAATCTCAGAGTTCTTTGAAGGTTGGAGAAAACAATAGAAATCTGGAAACA	480
Qy	247	VallilePheCysAspValLeuProGlyLysGluSerLeuGlyTyrPheValGlyValAsp	266
Db	481	GTTATATTCTGTGATGTTTTGCCAGAAAAAGAAGCTTAGATATTTTGTGGTGTGCAC	540
Qy	267	MetAspAsnProIleGlyAsnTrpAspGlyArgPheAspGlyVal---LeuCysSerPhe	285
Db	541	ATGGATAAACCTATTGGCAACTGGGATGGAAGATTTGATGGAGTGCAGCTTTGTAGTTTT	600
Qy	286	AlaCysValGluSerThrIleLeuLeuHisIleAsnAspIleIleProGluSerValThr	305
Db	601	CGGTGTGTTGAAGATACAAATTTATGTCACATCAATGATATCATCCCCAGAGAGTGTGACG	660
Qy	306	GlnGluArgArgProProLysLeuAlaPheMetSerArgGlyValGlyAspLysGlySer	325
Db	661	CAGGAAAGAGGCCCTCCCAGAACTTGCCTTTATGTCAAGAGGTGTGGGACAAAGGTTCA	720
Qy	326	SerSerHisAsnLysProLysAlaThrGlySerThrSerAspProGlyAsnArg---Arg	344
Db	721	TCCAGTCAATAAACCAGAAAGCTACAGAGTCTACCTCAGACCCCTGGAAAATAGAAACAGA	780
Qy	345	SerGluLeuPheTyrThrIleuasnGlySerSerValAspSerGlnProGlnSerIlysSer	364
Db	781	TCTGAATTAATTATACCTTAATATGGGTCTTCTGTGTGACTCACAAACCAACAATCCAAATCA	840
Qy	365	LysAsnThrTrpTyrIleAspGluValAlaGluAspProAlaLysSerLeuThrGluIle	384
Db	841	AAAAATACATGTCATTTGATGAAGTTCCAGAGACCCCTGC AAAATCTCTTACAGAGATA	900
Qy	385	SerThrAspPheAspArgSerSerProLeuGlnProProValAsnSerLeuThr	404
Db	901	TCTACAGACTTTTGACCGTTCTTCCACCACCACTCCAGCCCTCCTCCTGTGAACTCACTGACC	960
Qy	405	ThrGluAsnArgPheHisSerLeuP-oPheSerLeuThrLysMetProAsnThrAsnGly	424
Db	961	ACCGAGAACAGATTCCTCTTTTACCAVATCGAGTCTCACCAAGATGCCCAATACCAATGGA	1020
Qy	425	SerIleGlyHisSerProLeuSerLeuSerAlaGlnSerValMetGluGluLeuAsnThr	444
Db	1021	AGTATTGGCCACAGTCCACTTTCTCTGTGAGCCCACTCTGTAATGGAAGAGCTAAACACT	1080
Qy	445	AlaProValGlnGluSerProProLeuAlaMetProProGlyAsnSerHisGlyLeuGlu	464
Db	1081	GCACCCGTCAGAGAGTCCACCTTGGCCATGCCCTCCTGGGAACCTCATGGTCTAGAA	1140
Qy	465	ValGlySerLeuAlaGluValLysGluAsnProPropheTyrGlyValIleargTrpIle	484
Db	1141	GTGGGCTCATTTGGCTCAAGTTAAGGAGAACCCCTCCTCTTATGGGGTAAATCCGTGGATC	1200
Qy	485	GlyGlnProProGlyLeuAsnGluValLeuAlaGlyLeuGluLeuGluAspGluCysAla	504
Db	1201	GGTCAGCCACCAAGAGCTGAATGAAGTGTCTGCTGGACTGGAACTGGAAAGATGAGTGTCA	1260
Qy	505	GlyCysThrAspGlyThrPheArgGlyThrArgTyrPheThrCysAlaLeuLysLysAla	524
Db	1261	GGCTGTACGGATGGAACCTTCAGAGGCACTCGGTATTTTCCCTGTGCTGAGGAGCGC	1320
Qy	525	LeuPheValLysLeuLysSerCysArgProAspSerArgPheAlaSerLeuGlnProVal	544
Db	1321	CTGTTTGTGAAACTGAAGAGCTGCAGGCCGTGACTCTAGGTGTTGCAATCATTCAGCCGGTT	1380
Qy	545	SerAsnGlnIleGluArgCysAsnSerLeuAlaPheGlyGlyTyrLeuSerGluValVal	564
Db	1381	TCCAATCAGATTGAGCGCTGTAACTCTTTAGCAATTTGGAGGCTACTTAAAGTGAAGTAGTA	1440
Qy	565	GluGluAsnThrProProLysMetGluLysGluGlyLeuGluIleMetIleGlyLysLys	584

```
Db      1 GTTGTAGAGATAAATGAAAGTTCCACAGAGTTACTTTTGGCAATTTACCAATTTGTGAGGAG 60
Qy      ArgPheSerLeuPheLeuAsnArgLeuSerLysGlyLeuGlnIleAspValGly 127
Db      61 AGGTTACAGCTGTTTAAACACAGAAACAGACTAGTAAGGCTCCAAATAGAGCTGGGC 120
Qy      128 CysProValLysValGlnLeuArgSerGlyGluGlyLysPheProGlyValValArgPhe 147
Db      121 TGTCTCTGAAAGTACAGCTGAGACTCTGGGGAAGAAAAATTTCTCGAGTTGTACGCTTC 180
Qy      148 ArgGlyProLeuLeuAlaGluArgThrValSerGlyIlePhePheGlyValGluLeuLeu 167
Db      181 AGAGGACCCCTGTTAGCAGAGAGACAGTCTCCGGAATATTCTTTGGAGTTGAATTTGCTG 240
Qy      168 GluGluGlyArgGlyGlnGlyPheThrAspGlyValTyrGlnGlyLysGlnLeuPheGln 187
Db      241 GAAGNAGTCTGTTGCTCAAGGTTTCACTGACGGGTGTACNAGGGAACAGACTTTTTCAG 300
Qy      188 CysAspGluAspCysGly---PheValAlaLeuAspLysLeuGluLeuIleGluAspAsp 206
Db      301 TGTGATGAAGATTGTGGCGTGTGTTTGGCATTTGCAATGGACAAGCTAGAACTCATAGAAGATGAT 360
Qy      207 AspThrAlaLeuGluSerAspTyrAlaGlyProGlyAspThrMetGlnValGluLeuPro 226
Db      361 GACACTGCATTGGAAGTGATTAACGACAGTCTCGGAGACCAATGCAAGGTGCAACTTCCT 420
Qy      227 ProLeuGluIleAsnSerArgValSerLeuLysGlyGlyGluThrIleGluSerGlyThr 246
Db      421 CCTTTGGAAATAAAGCTCCAGAGTTTCTTTGAAGGTTGGGAAACAAATAGAACTCTGGAACA 480
Qy      247 ValIlePheCysAspValLeuProGlyLysGluSerLeuGlyTyrPheValGlyValAsp 266
Db      481 GTTATATTCTGTGATGTTTTCGCCAGGAAAGAAAGCTTAGGATATTTTGTGTGTGGAC 540
Qy      267 MetAspAsnProIleGlyAsnTrpAspGlyArgPheAspGlyVal---LeuCysSerPhe 285
Db      541 ATGGATAAACCTATTGGCAACTGGGATGGAAGATTTTGATGGAGTGCAGCTTTGTAGTTT 600
Qy      286 AlaCysValGluSerThrIleLeuLeuHisIleAsnAspIleIleProGluSerValThr 305
Db      601 GCCTGTGTTGAAAGTACAATTCATTGTCACATCAATGATATCATCCACAGAGTGTGACG 660
Qy      306 GlnGluArgArgProProLysLeuAlaPheMetSerArgGlyValGlyAspLysGlySer 325
Db      661 CAGGAAGAGGGGCTCCCAAACTTGCTTTATGTCAAGAGTGTGTGGGACAAAGGTTC 720
Qy      326 SerSerHisAsnLysProLysAlaThrGlySerThrSerAspProGlyAsnArg---Arg 344
Db      721 TCCAGTCATAATAAACCAAGGCTACAGGATCTACCTCAGACCTCGAAATAGAAACAGA 780
Qy      345 SerGluLeuPheTyrThrLeuAsnGlySerSerValAspSerGlnProGlnSerLysSer 364
Db      781 TCTGAAATATTTTATACCTTTAAATGGGTCTTCTGTTGACTCACAAACCAATCCAAATCA 840
Qy      365 LysAsnThrTrpTyrIleAspGluValAlaGluAspProAlaLysSerLeuThrGluIle 384
Db      841 AAAAAATACATGGTACATTGATGAAGTTGCAAGAGACCTCGCAAAATCTTTACAGAGATA 900
Qy      385 SerThrAspPheAspArgSerProProLeuGlnProProValAsnSerLeuThr 404
Db      901 TCTACAGACTTTGACCGTCTTCCACCACTCCAGCTCCTCGCTGTGAACTCACTGACC 960
Qy      405 ThrGluAsnArgPheHisSerLeuProPheSerLeuThrLysMetProAsnThrAsnGly 424
Db      961 ACCGAGAACAGATTCACCTCTTTACCATTCAGTCTCACCAGATGCCCAATACCAATGGA 1020
Qy      425 SerIleGlyHisSerProLeuSerLeuSerAlaGlnSerValMetGluGluLeuAsnThr 444
Db      1021 AGTATTTGGCCACAGTCCACTTTCTCTGTGACCCAGTCTGTAATGGAAGAGCTAAACACT 1080
Qy      445 AlaProValGlnGluSerProProLeuAlaMetProProGlyAsnSerHisGlyLeuGlu
```

```
Db      1081 GCACCCCTCCAAGAGAGTCCACCCCTTGGCCATGCTCTCTGGGAACCTCATGCTGTAGAA 1140
Qy      465 ValGlySerLeuAlaGluValLysGluAsnProProPheTyrGlyValIleArgTrpIle 484
Db      1141 GTGGGCTCATTTGGCTGAAGTTAAGGAGAACCTCTCTTCTATATGGGTTAATCCGTTGGATC 1200
Qy      485 GlyGlnProProGlyLeuAsnGluValLeuAlaGlyLeuGluLeuGluAspGluCysAla 504
Db      1201 GGTCAAGCACACAGGACTGAATGAAGTGTCTCGCTGGACTGGAACCTGGAAGATGAGTGTGCA 1260
Qy      505 GlyCysThrAspGlyThrPheArgGlyThrArgTyrPheThrCysAlaLeuLysLysAla 524
Db      1261 GGCTGTACGATGAACCTTCAGAGGCACTCGGTATTTTTCACCTGTGCTGCTGAAAGAGCG 1320
Qy      525 LeuPheValLysLeuLysSerCysArgProAspSerArgPheAlaSerLeuGlnProVal 544
Db      1321 CTGTTTGTGAACATGAAGAGCTGCAGGCTGACTCTAGGTTTGCATCATTCGACGCGGT 1380
Qy      545 SerAsnGlnIleGluArgCysAsnSerLeuAlaPheGlyGlyTyrLeuSerGluValVal 564
Db      1381 TCCAATCAGATTGAGCGCTGTAACTCTTTAGCATTTGGAGGCTACTTAAGTGAAGTAGTA 1440
Qy      565 GluGluAsnThrProProLysMetGluLysGluGlyLeuGluIleMetIleGlyLysLys 584
Db      1441 GAAAGAAAATACTCCACAAAATAAGAAAAGAGGCTTGGAGATAAATGATTGGGAAGAG 1500
Qy      585 LysGlyIleGlnGlyHisTyrAsnSerCysTyrLeuAspSerThrLeuPheCysLeuPhe 604
Db      1501 AAAGGCATCCAGGCTCATTAATCTTGTGTACTTAGACTCAACCTTATTTCTGCTTATTT 1560
Qy      605 AlaPheSerSerValLeuAspThrValLeuLeuArgProLysGluLysAsnAspValGlu 624
Db      1561 GCITTTAGTTCTGTTCTGGACACTGTGTACTTAGACCCAAAGAAAGACGATGTAGAA 1620
Qy      625 TyrTyrSerGluThrGlnGluLeuLeuArgThrGluIleValAsnProLeuArgIleTyr 644
Db      1621 TATTATAGTGAACCCCAAGAGCTACTGAGGACAGAAATTTGTTAATCTCTGAGAATATAT 1680
Qy      645 GlyTyrValCysAlaThrLysIleMetLysLeuArgLysIleLeuGluLysValGluAla 664
Db      1681 GGATATGTGTGTGCCCAAAAATTTATGAACTGAGGAAAATACTTTGAAAAGGTGGAGGCT 1740
Qy      665 AlaSerGlyPheThrSerGluLysAspProGluGluPheLeuAsnIle 681
Db      1741 GCATCAGATTATCTCTGAGAAAAGGT-----GACCATCTTAACTTA 1785
```

Search completed: September 18, 2004, 19:01:03
Job time : 8564 secs